NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

ASSESSING THE EFFECTIVENESS OF CUMULATIVE SUM NORMAL- AND POISSON-BASED TESTS FOR DETECTING RARE DISEASES

by

Manuel E. Ganuza

December 2010

Thesis Advisor:	Ronald D. Fricker, Jr.
Second Reader:	Krista D. Hanni

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** December 2010	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE Assessing the Effectiveness of Cumulative Sum Normal- and Poisson-Based Tests for Detecting Rare Diseases		**5. FUNDING NUMBERS**	
6. AUTHOR(S) Manuel E. Ganuza			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000		**8. PERFORMING ORGANIZATION REPORT NUMBER**	
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) Defense Threat Reduction Agency		**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**	
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number N.A.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited		**12b. DISTRIBUTION CODE**	

13. ABSTRACT (maximum 200 words)

The early detection of a rare disease outbreak is of vital importance in public health and national defense. The six Category A biological agents designated by the Centers for Disease Control and Prevention are causal agents of rare diseases. The *Francisella tularensis* is one of these and is the causal agent of the tularemia disease. Tularemia is used as the motivating problem to evaluate and compare the effectiveness of the normal and Poisson-based CUSUM in the early detection of an outbreak using simulated rare disease data based on its theoretical behavior at varying occurrence outbreak means.

In this study, a mean relationship between the nonoutbreak λ_0 and outbreak λ_1 means was found, $\lambda_1 = 7.40\ \lambda_0$. Simulations were run to study the mean relationship in three cases: one theoretical case, where the normal- and Poisson-base CUSUM are exactly equal; and two other extreme cases. The computational results show that when λ_1 is very close to λ_0 the normal-based CUSUM system behaves improperly resulting in early detection delays, when λ_1 is equal or greater than λ_0 the normal- and Poisson-based CUSUM behave almost equally if the threshold for the normal-based CUSUM is selected properly. Methods to determined proper thresholds are also given.

14. SUBJECT TERMS Biosurveillance, Rare Disease, Tularemia, Cumulative Sum, CUSUM			**15. NUMBER OF PAGES** 79
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

ASSESSING THE EFFECTIVENESS OF CUMULATIVE SUM NORMAL- AND POISSON-BASED TESTS FOR DETECTING RARE DISEASES

Manuel E. Ganuza
Lieutenant Commander, United States Navy
B.S., Instituto Universitario Politecnico de las Fuerzas Armadas Nacionales de Venezuela, 1991

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
December 2010

Author: Manuel E. Ganuza

Approved by: Ronald D. Fricker, Jr., PhD
Thesis Advisor

Krista D. Hanni, PhD
Second Reader

Robert F. Dell, PhD
Chairman, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The early detection of a rare disease outbreak is of vital importance in public health and national defense. The six Category A biological agents designated by the Centers for Disease Control and Prevention are causal agents of rare diseases. The *Francisella tularensis* is one of these, and is the causal agent of the tularemia disease. Tularemia is used as the motivating problem to evaluate and compare the effectiveness of the normal and Poisson-based CUSUM in the early detection of an outbreak, using simulated rare disease data based on its theoretical behavior at varying occurrence outbreak means.

In this study, a mean relationship between the nonoutbreak λ_0 and outbreak λ_1 means was found, $\lambda_1 = 7.40\ \lambda_0$. Simulations were run to study the mean relationship in three cases: one theoretical case, where the normal- and Poisson-base CUSUM are exactly equal; and two other extreme cases. The computational results show that when λ_1 is very close to λ_0 the normal-based CUSUM system behaves improperly, resulting in early detection delays; when λ_1 is equal or greater than λ_0 the normal- and Poisson-based CUSUM behave almost equally if the threshold for the normal-based CUSUM is selected properly. Methods to determined proper thresholds are also given.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ATFS	Average Time to False/First Signal
BARDA	Biomedical Advance Research and Development Authority
CDC	Centers for Disease Control and Prevention
CIDRAP	Center for Infectious Disease Research And Policy
CUSUM	Cumulative Sum
DHHS	United States Department of Health and Human Services
F. tularensis	*Francisella tularensis*
GI	Gastrointestinal Illness
h	Threshold
HSPD-21	Homeland Security Presidential Directive 21
iid	independent and identically distributed
ILI	Influenza-Like Illness
Norm	Normal Distribution Probability Density Function
Pois	Poisson Distribution Probability Mass Function
Sys	System
VHF	Viral Hemorrhagic Fevers

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

Biosurveillance is a critical component of public health as well as national defense and homeland security programs. In particular, early detection of naturally-occurring diseases and bioterrorist attacks—in human and animal populations, food, water, agriculture, and the environment in general—is essential for timely response in order to mitigate and reduce the consequences of such an outbreak or attack.

To address these threats, the Department of Defense, as well as the Department of Human Health and Services, have implemented various *biosurveillance systems.* These systems rely on statistical algorithms for disease early detection. Many of the algorithms in use implicitly or explicitly assume the data is normally distributed; however, most syndromic surveillance data is discrete (usually daily counts of syndromes or disease) and thus, by definition, are not normally distributed.

This thesis compares the performance of two variants of the cumulative sum (CUSUM) algorithm, one based on the normal distribution, and the other based on the Poisson distribution. The latter is likely to be more appropriate when monitoring for rare diseases. The comparison is conducted assuming Tularemia is the agent being monitored by the biosurveillance system. Tularemia is listed as a Category A agent by the Centers for Disease Control and Prevention because of its virulence and capability for weaponization. However, Tularemia also occurs naturally, though very rarely, with about 125 diagnosed cases per year in the United States.

The results of this research show that (incorrectly) monitoring a rare disease such as Tularemia with a CUSUM based on the normal distribution results in either an excessive number of false positive signals or a delay in detecting an outbreak (in comparison to a CUSUM based on the Poisson distribution). The delays observed in this research were from 5 to 15 days. Thus, to monitor a rare disease such as Tularemia, a Poisson-based CUSUM should be included in biosurveillance systems.

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

Ante todo, doy gracias a Dios y a La Virgen de Guadalupe, a María Lionza, Guaicaipuro, Negro Primero y Los Hermanos.

Thank you, Father, Mother, and Brother for your support in the good and not so good times. You are the pillars from which I have grown.

Professor Ron Fricker, muchas gracias por toda su paciencia. From the first day of class, you were able to bring into me the curiosity and thirst to learn more and apply the mathematical concepts into reality.

Special gratitude goes to the United States Navy for giving me the opportunity to grow academically, to the Naval Postgraduate School and to the faculty of the Operations Research department, for having the patience and dedication to guide me through the path of applied mathematics.

THIS PAGE INTENTIONALLY LEFT BLANK

I. RARE DISEASES IN PUBLIC HEALTH AND NATIONAL DEFENSE

Schoenbach and Rosamond (2000) define a disease as a definite pathological process having a characteristic set of signs and symptoms. The disease etiology (i.e., cause), pathology (i.e., manifested symptoms), and prognosis (i.e., expected outcome) may be known or unknown. In epidemiology, diseases are thought of as processes where the natural history of disease describes the uninterrupted progression in a host of the disease's biological development from the moment it is initiated by causal agent exposure. Infectious diseases are transmissible from human-to-human or animal-to-human, spread geographically and or increase the size of the infected population due to the capability of being communicable by infection (Dorland's Medical Dictionary, 2007). Infectious diseases may be due to organisms ranging in size from viruses to parasitic worms, and which act as causal agents that get inside a host, human or animal and cause disease.

The study of infectious disease has unique challenges. Infectious diseases that naturally occur as epidemic diseases are easy to study as each occurrence represents a form of natural experiment and provides a contrast between before and after. In the case of an endemic infectious disease,[1] there is no obvious contrast to stimulate perception of new events or new modes of living that could have introduced the disease. Rare and new diseases are more difficult because there are very few events to study (Schoenbach & Rosamond, 2000). These rare and new infectious diseases are called emerging infectious diseases.

Morse (1995) gave a formal and accepted definition for an emerging infectious disease: "infections that have newly appeared in a population or have existed but are rapidly increasing in incidence or geographic range." Morse (1995) stresses the importance of recognizing the abundance of new infectious diseases with zoonoses (i.e., animal) origin "the zoonotic pool—introductions of infections from other species to

[1] A disease is said to be endemic in a population when the disease naturally occurs in that population without the need for external exposure.

humans—is an important and potentially rich source of emerging diseases; periodic discoveries of new zoonoses suggest that the zoonotic pool appears by no means exhausted."

The early detection of emerging infectious diseases is of vital concern in both public health and national defense. Public health has historically been concerned with the control and eradication of infectious diseases in order to prevent or minimize the impact of natural disease outbreaks on the general population. In contrast, national (and homeland) defense has been more concerned with the use of biological agents by adversaries as a means of inflicting harm, terror, and economic disruption on the general population, military forces, or both. Surveillance and detection of biological agents, either naturally occurring or man-made, is where public health and national defense/homeland security interests converge. The Centers for Disease Control and Prevention (CDC) has designated six biological agents, including pathogens that are rarely seen in the United States, as Category A agents. These are high-priority agents that include organisms that pose a risk to national security because they can be easily disseminated or transmitted from person to person, result in high mortality rates, have the potential for major public health impact, might cause public panic and social disruption and/or require special action for public health preparedness (CDC, 2010).

Table 1 lists the six Category A biological agents. Inglesby et al. (1999, 2002) studied anthrax; *Clostridium botulinum* toxin was studied by Arnon et al. (2001); Inglesby et al. (2000) studied the plague; samllpox was studied by Henderson et al. (1999); Dennis et al. (2001) studied tularemia and Borio et al. (2002) studied hemorrhagic fever viruses.

Disease Common Name	Biological Agent Name	Remarks
Anthrax	*Bacillius anthracis*	Bacteria, Inglesby et al. (1999) and Inglesby et al. (2002)
Botulism	*Clostridium botulinum*	Bacteria toxin, Arnon et al. (2001)
Plague	*Yersinia pestis*	Bacteria, Inglesby et al. (2000)
Smallpox	Smallpox	Genus *orthopoxvirus*, Henderson et al. (1999)
Tularemia	*Francisella tularensis*	Bacteria, Dennis et al. (2001)
Viral Hemorrhagic Fevers (VHF)	Filoviridae Arenaviridae Bunyaviridae Flaviviridae	Four distinct family viruses, Borio et al. (2002)

Table 1. List of Category A Biological Agents (CDC, 2010)

A. SURVEILLANCE AND DETECTION

Biosurveillance is a critical component of public health and homeland security programs, where early warning, detection, and recognition of infectious diseases—in human and animal populations, food, water, agriculture, and the environment in general—is essential for timely response in order to mitigate and reduce the consequences of a disease outbreak. Traditional biosurveillance and related epidemiological methods, such as routine analyses of birth and death or infectious disease reports, do not achieve the necessary and desired speed of detection. In an effort to achieve the requisite of early detection capability, these traditional health methods have expanded to include the surveillance and analysis of non-specific health and health-related data. The idea is that these types of data might provide advance indicators of disease activity and threats to human or animal health (Fricker, 2010a)

Epidemiology, as a multidisciplinary science, studies health and disease in human populations. Traditionally, epidemiology is an applied research discipline focused on understanding the causes and effects of disease. Over time, epidemiology has diversified into many forms and one of particular interest is *epidemiologic surveillance*. See Schoenbach and Rosamond (2000); and Bohpal (2002). Homeland Security Presidential Directive 21 (HSPD-21) defines epidemiologic surveillance as "the process of actively gathering and analyzing data related to human health and disease in a population in order

to obtain early warning of human health events, rapid characterization of human disease events, and overall situational awareness of disease activity in the human population." Epidemiologic surveillance is therefore a subset of biosurveillance focused on human populations. *Syndromic surveillance* is a subset of epidemiologic surveillance that uses non-specific health data to identify disease outbreaks as early as possible. See Fricker (2010a, 2010b, 2010c, 2008) for additional information about biosurveillance, epidemiologic surveillance, and syndromic surveillance.

According to Schoenbach and Rosamond (2002), in addition to the symptomatic period, a disease has two important periods, the presymptomatic and the postmorbid. The postmorbid period begins at the end of disease pathological and clinical course, in other words it is the aftermath of the disease (Seiger, 1961). The presymptomatic period is the time between infection and the clinical manifestations of the disease occur. For infectious diseases, there are two important events in the presymptomatic period, disease detection and the onset of infectiousness (Schoenbach & Rosamond, 2000). The ideal syndromic surveillance system would detect an infectious disease when it is presymptomatic and prior to the onset of infectiousness (Fricker, 2010a).

As previously mentioned, syndromic surveillance is a more specific type of epidemiological surveillance, with the fundamental objective of identifying illness clusters early, before diagnoses are confirmed and reported to public health agencies. The goal is to mobilize a response as early and rapidly as possible in order to reduce morbidity (i.e., disease incidence) and mortality. As Henning (2003) points out, the ability of syndromic surveillance to detect outbreaks earlier than conventional surveillance methods depends on such factors as the size of the outbreak, the population dispersion of those affected, data sources and syndrome definitions used, surrogate data sources, non-specific disease syndromes, the criteria for investigating threshold alerts, and the health-care provider's ability to detect and report unusual cases..

B. ISSUES IN BIOSURVEILLANCE

An epidemic, endemic, or other large outbreak of an emerging infectious disease can be the result of natural causes, warfare, or terrorism (i.e., it can be either naturally occurring or intentional). At the initial stages of the outbreak, the causes of the outbreak may be unclear. However, as just described, the early detection of a disease in a population is of interest to public health and national defense officials so that measures can be taken to mitigate the effect or effects of the outbreak.

Biosurveillance systems rely on statistical algorithms for disease early detection. Their use presents various challenges related to data availability, data quality, algorithm appropriateness and effectiveness, and other factors. See Fricker (2010b), Fricker and Rolka (2006) and Shmueli and Burkom (2010) for additional discussion. In terms of algorithm appropriateness, many of the algorithms in use implicitly or explicitly assume the data is normally distributed; however, most syndromic surveillance data is discrete (usually daily counts of syndromes or disease) and thus by definition are not normally distributed. This limitation is sometimes overcome by monitoring the residuals of a model used to account for systematic effects in the data. For further discussion, see Hagen (2010); Fricker, Hegler and Dunfee (2008); and Fricker, Knitt and Hu (2008).

However, there are situations in which the use of an algorithm that assumes normality is inappropriate. For example, naturally occurring rare diseases will not follow the normal (Gaussian) distribution, nor will the residuals from a model likely be normally distributed. However, numerous biosurveillance systems blindly use statistical algorithms based in the normal distribution without assessing whether the data to which they are applying it meet this assumption.

The use of inappropriate algorithms in a biosurveillance system may cause excessive numbers of false signals of an outbreak that results in: (1) unnecessary use of limited medical and public health resources and (2) a loss of confidence in the biosurveillance system. Inappropriate use may also delay the detection of an outbreak,

which gives more time for the emerging infectious or rare disease to spread within the population and spread out geographically, both making the disease more difficult to control and increasing its economic impact.

C. MOTIVATING PROBLEM

This thesis assesses the performance of a particular statistical algorithm for detecting a rare disease. Because the disease is rare, the daily counts of occurrence are usually zero and only during an outbreak do the counts tend to be positive.

For this study, tularemia was chosen as the disease of interest. Tularemia is of zoonotic origin and is caused by a bacterium named *Francisella tularensis* (*F tularensis*). It is naturally occurring and is one of the most infectious pathogenic bacteria known, requiring inoculation or inhalation of just 10 to 50 organisms of the most virulent and infectious subspecies to cause disease. Humans become incidentally infected through diverse environmental exposures and can develop severe and sometimes fatal illness with a 30% to 60% mortality rate if left untreated. There is no human-to-human transmission, and it is one of the pathogens monitored by the BioWatch aerosol surveillance program for potential bioterrorist attacks (Barns, Grow, Okinaka, Keim, & Kuske, 2005). The Centers for Disease Control and Prevention has designated *F. tularensis* a Category A agent for it extreme infectivity, ease of dissemination, and substantial capacity to cause illness and death (CDC, 2010).

Tularemia research continues to be of importance in public health and national defense. For example, the Biomedical Advanced Research and Development Authority (BARDA) announced the decision from the U.S. Department of Health and Human Services (DHHS) to invest a total of $64 million to further the development of a new broad spectrum antibiotic for treating plague, tularemia, and various resistance infections (BARDA, 2010).

II. TULAREMIA

Tularemia is colloquially known as rabbit fever or deer fly fever. Tularemia was first identified in 1911, during an outbreak investigation in Tulare County, California, for which it has been named. The *Francisella tularensis* bacterium has various characteristics that make it effective for development and use as a biological weapon, including the virulence, infectiousness, and genetic manipulation (Davis, 1999; Dennis et al., 2001), as well as its ability to survive in mud or water for at least a year (Barns et al., 2005). The Center for Infectious Disease Research and Policy (CIDRAP), http://www.cidrap.umn.edu, at the University of Minnesota, provides a large number of references and information related to infectious diseases and those biological agents with possible use in bioterrorism.

A. SUBSPECIES

F. tularensis subspecies can be differentiated by biochemical and molecular tests. The four well known subspecies are *F tularensis* subs *tularensis* (Jellison type A), *F tularensis* subs *holarctica* (Jellison type B), *F tularensis* subs *mediaasiatica* and *F. tularensis* subs *novicida* (CIDRAP, 2010)

The most virulent of the known subspecies is Jellison type A, found in North America, with a fatality rate of up to 30% before the introduction of antibiotic regimens, and a low fatality rate with the appropriate antibiotic therapy. Jellison type B is less virulent and is the prevalent form of tularemia, found in North America and elsewhere, with a low fatality rate even without treatment. *F. tularensis* subs *novicida* (the mildest presentation) is the rarest form of tularemia and is typically associated with waterborne-acquired infection (Chu, Mead, & Sjöstedt, 2010; Sjöstedt, 2010). Jellison type A incubation period is 3 to 5 days (range 1 to 14 days) after inoculation and it is the more likely strain to be used as a biological agent (CIDRAP, 2010).

The clinical presentation of symptoms may be confused with those of the plague and other infectious diseases such as staphylococcal and streptococcal infections, cat-scratch fever, and tuberculosis (Chu, Mead, & Sjöstedt, 2010; Sjöstedt, 2010).

B. EPIDEMIOLOGY

1. Modes of Transmission

As summarized in Table 2, the mode of transmission into humans is via a group of reservoirs,[2] vectors[3] and environmental mediums that have a very wide range of forms. The primary reservoirs are small- to medium-size mammals, such as rabbits, and the primary documented vectors are ticks and biting flies (CIDRAP, 2010). Other less common reservoirs includes a variety of vertebrate (some species of birds and fish) and invertebrate taxa (Dudley, 2010). The period of communicability varies between vectors; flies can be infective for up to 14 days and ticks infective throughout their lifetime (Chu, Mead, & Sjöstedt, 2010; Sjöstedt, 2010).

Modes of Transmission
Bites by infected arthropods
Direct contact with infected animals
Handling of infectious animal tissue or fluids
Ingestion of contaminated food, water or soil
Inhalation of infectious aerosols, including dust from contaminated hay and aerosols generated by lawn mowing and brush cutting
Exposure in the laboratory setting
Possibly direct contact with contaminated soil or water

Table 2. Modes of Tularemia Transmission to Humans (CIDRAP, 2010)

[2] A reservoir is an alternate or passive host or carrier that harbors pathogenic organisms or parasites without injury to itself, and serves as a source from which other individuals can be infected (Dorland's Medical Dictionary, 2007).

[3] A vector is a carrier, especially an animal, such as an arthropod, that transfers an infective agent from one host to another (Dorland's Medical Dictionary, 2007).

2. Incidence

The geographic distribution of tularemia has included the entire continental United States (Chu, Mead, & Sjöstedt, 2010; Sjöstedt, 2010). The CDC's Morbidity and Mortality Weekly Report (MMWR) reported an average of 124 cases (range 86–193) per year from 1990 to 2000, with the highest number of cases in counties throughout Arkansas and Missouri, eastern Oklahoma and Kansas, southern South Dakota and Montana, and Massachusetts (CDC, 2002). Between 2000 and 2008, the average number of cases was 125 per year, with the highest incidence in Missouri, Arkansas, Oklahoma, and Massachusetts. A particular area in Massachusetts is Martha's Vineyard, where the majority of those cases occurred (Weinberg & Branda, 2010). Figure 1 shows the geographical distribution of reported cases of Tularemia in the United States between 2000 and 2008 (Dudley, 2010).

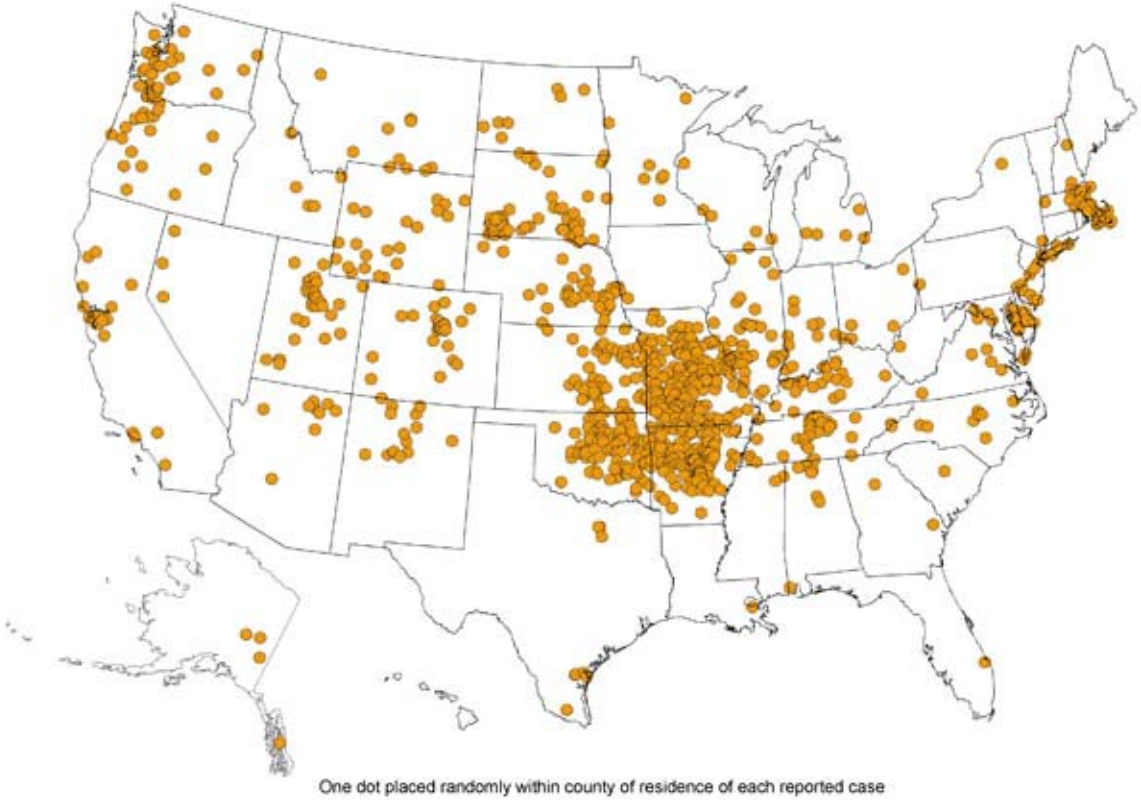

Figure 1. Locations of Reported Cases of Tularemia in the United States, 2000–2008 (Dudley, 2010, p. 10)

C. CLINICAL SYNDROMES

A clinical syndrome is the manifestation of a disease process in a patient pertaining to or founded on actual observation, treatment, and laboratory findings. The tularemia disease manifests on a patient in various clinical forms depending in the virulence of the strain, dose, and inoculum (Dennis et al., 2001).

1. Glandular and Ulceroglandular Tularemia

Organisms enter the host through apparent breaks in the skin surface due to the bite of infective arthropods, direct contact with infectious materials, or the puncture of skin with a sharp object. The infection dose for humans is 10 to 50 organisms. From the site of inoculation, the organisms spread to the regional lymph nodes (CIDRAP, 2010). The onset is 3 to 6 (range 1 to 14) days after exposure showing symptoms of chills, fever, head and muscle pain, and prostration (Foley & Nieto, 2010).

2. Pneumonic Tularemia

Infection happens through the lungs either via inhalation or through dissemination in the bloodstream. The infectious dose is 10 to 50 organisms. The organisms rapidly, entering the pulmonary area (within minutes) and begin replicating. This explosive capacity to replicate appears to be an important factor in virulence associated with pulmonary infection (CIDAP, 2010). Ulcers may be absent; patients may have a dry cough, shortness of breath, and chest pain, with patchy infiltrates (fluids), lobar pneumonia,[4] or bloody pleural effusion.[5] The fatality rate can be up to 40% without antibiotic treatment (Foley & Nieto, 2010).

[4] A form of pneumonia that affects the lobe of a lung.

[5] Excessive bloody fluids seeping in the chest cavity.

3. Oculoglandular Tularemia

Organisms enter via ocular/conjunctiva, with unilateral conjunctival ulcers, purulent[6] conjunctivitis, periorbital[7] edema[8] and enlargement of the cervical[9] and preauricular[10] lymph nodes[11] (Foley & Nieto, 2010).

4. Oropharyngeal and Gastrointestinal Tularemia

Organisms enter via the mucous membrane of the oropharynx following ingestion of infected material (meat or water). Lesions develop primarily in tonsillitis and pharyngitis[12] and lymph node; throughout the gastrointestinal (GI) tract there may be a few lesions. The usual symptoms are sore throat and or abdominal pain, with sporadic vomiting, diarrhea, and GI bleeding. The fatality rate can be up to 60% (Foley & Nieto, 2010).

5. Typhodial Tularemia

Typhoidal tularemia is secondary to pneumonic tularemia and occasionally occurs after ulceroglandular tularemia. The characteristic symptoms of typhoidal tularemia are not observed and there is no visible site of inoculation (Foley & Nieto, 2010).

D. TULAREMIA IN BIOLOGICAL WARFARE

The use of *F tularensis* in warfare can be traced to 14th century BC, where the Hittites of Anatolia sent infected animals and people into western Anatolia in order to cause outbreaks inside enemy territory further spreading an already occurring tularemia epidemic that lasted 35 to 40 years (Trevisanato, 2007).

---

[6] Conjunctivitis containing pus.

[7] The periosteum of the bones forming the orbit, eye socket.

[8] Accumulation of excess fluids.

[9] Pertaining to the neck.

[10] In front of the aurice of the ear.

[11] Filter the lymphatic fluid. Located in different regions of the body. Main regions are the neck, underarms, upper chest, and the groin (MedicineNet, n.d.)

[12] Inflammation of the throat.

In the modern era, *F tularensis* has been weaponized by freeze drying a bacteria-laden slurry and milling it into a fine powder for aerosol delivery (Cronquist, 2004). The Japanese germ warfare research units studied the potential of *F tularensis* as a modern biological weapon while operating in Manchuria between 1932 and 1945 (Cronquist, 2004). The use of *F tularensis* as a biological weapon has been suggested to be the cause of a Tularemia outbreak affecting tens of thousand Soviet and German soldiers during the battle of Stalingrad in World War II; however, it is also thought to be the result of natural causes (Croddy, 2001). After World War II, further research was conducted to understand the pathophysiology of tularemia and to develop vaccines and antibiotic prophylaxis and treatment regimens (Dennis et al., 2001).

In the late 1960s, *F tularensis* was one of several biological weapons stockpiled for delivery as aerosols, and then, in 1969, President Nixon gave the declaration of biological disarmament. The Biological and Toxin Weapons Convention of 1972 placed the possible use of biological weapons very low. There are thoughts that the former Soviet Union continued with the research and development of biological weapons under an organization named Chief Directorate for Biological Preparations (mostly known as Biopreparat). Biopreparat continued operation until the early 1990 with the production of *F. tularensis* strains engineered to be resistant to antibiotics and vaccines (Davis, 1999).

By the late 1990s to early 2000s, the CDC studied the economic impact of a bioterrorist attack based on the model used by the World Health Organization to estimate the impact of an intentional aerosol dispersal of 50 kg of virulent *F tularensis* over a metropolitan area with 5 million inhabitants. The economic impact is estimated at $5.4 billion for every 100,000 persons exposed. The estimated number of persons exposed is 250,000 incapacitating casualties, including 19,000 deaths; the number could be greater if a genetically modify strain is used (Dennis et al., 2001).

E. RECENT NATURAL OUTBREAKS

Tularemia disease is currently a rare disease in the United States, with a peak number of 2,300 cases reported in 1939 prior to development of antibiotic therapy. The

only two pneumonic tularemia outbreaks in the United States have occurred in Martha's Vineyard, Massachusetts, in 1978 and 2000 (Hornick, 2001).

The Martha's Vineyard 2000 pneumonic tularemia outbreak resulted in 15 patients, of which 11 cases where pneumonic tularemia. The strain was identified to be *F tularensis* type A from samples of the only fatality. The initial reports started with five cases in July 2000 and by late August, six additional cases were identified. The modes of transmission were associated with lawn mowing and brush cutting (Feldman et al., 2001).

In the fall of 2005 in Washington, District of Columbia, traces of *F tularensis* were detected on the Capitol Mall by air sensors under the federal BioWatch program. Fortunately, no cases of tularemia were reported among the people participating in an antiwar demonstration and a National Book Festival (CIDRAP, 2005).

Worldwide, natural outbreaks have been reported from Scandinavia with an average of 3 to 4 cases per 100,000 people, Southern Europe, former Soviet Union, Japan, Canada and Mexico. In Spain, an outbreak with 585 cases is associated with exposure to infected hares. In Turkey and Bulgaria, large outbreaks were associated with direct exposure with contaminated drinking water. In Sweden and Finland, an epidemic of respiratory tularemia occurred after mosquito bites as well as hay contaminated from infected rodents and the bacteria became airborne while handling and moving hay (Foley & Nieto, 2010). In Kosovo, a large outbreak of 327 cases occurred after handling contaminated food and water from infected rodents (Reintjes et al., 2002).

THIS PAGE INTENTIONALLY LEFT BLANK

III. CUMULATIVE SUM

The Cumulative Sum (CUSUM) control chart (Page, 1954; Lorden, 1971) is a sequential hypothesis test for a change from a known, $\boldsymbol{f_0}$, nonoutbreak distribution to a known or estimated outbreak distribution, $\boldsymbol{f_1}$. The method defines a threshold h and monitors the statistic $\boldsymbol{C_t}$; it signals when the CUSUM statistic is greater or equal to the threshold, $C_t \geq h$.

The CUSUM satisfies the recursion

$$C_t = \max\left(0, C_{t-1} + \ln\frac{f_1(X_t)}{f_0(X_t)}\right) \tag{1}$$

where,

$\boldsymbol{C_t}$ is the current value of the CUSUM on day t

$\boldsymbol{C_{t-1}}$ is the value of the CUSUM on day t-1

$\boldsymbol{f_0}$ is the rare disease nonoutbreak distribution

$\boldsymbol{f_1}$ is the rare disease outbreak distribution

ln [$\boldsymbol{f_1(X_t)}$ / $\boldsymbol{f_0(X_t)}$] is the log-likelihood ratio for X_t relating the nonoutbreak distribution and the outbreak distribution. X_t is the observation for day t.

The method usually starts at $C_0 = 0$ and it stops and concludes that an outbreak may be occurring fat the first time when $C_t \geq h$. The implementation of the CUSUM requires choosing a threshold h to achieve a desired level of performance. Assuming that all observations are independent and identically distributed (*iid*) according to the nonoutbreak distribution, $\boldsymbol{f_0}$, the threshold h is chosen to make the average time between false signals (ATFS) sufficiently large. The ideal threshold has a large time between false signals when there is no outbreak and a small delay to the first real signal when there is an outbreak.

A. NORMAL-BASED CUSUM

When the nonoutbreak, $\boldsymbol{f_0}$, and outbreak, $\boldsymbol{f_1}$, distributions have normal densities with common variance $\boldsymbol{\sigma^2}$ and means $\boldsymbol{\mu_0}$ and $\boldsymbol{\mu_1} = \boldsymbol{\mu_0}+\boldsymbol{\delta\sigma}$, respectively, then Equation 1 reduces to

$$C_t = \max\left(0, C_{t-1} + X_t - \frac{\mu_1 - \mu_0}{2}\right) \tag{2}$$

This normal-based CUSUM will detect only increases in the mean, so it is a one-sided CUSUM.

B. POISSON-BASED CUSUM

When the nonoutbreak, $\boldsymbol{f_0}$, and outbreak, $\boldsymbol{f_1}$, distributions have Poisson mass functions with means λ_0 and λ_1, respectively, so that $\boldsymbol{f_0} = \boldsymbol{Pois}(\ \lambda_0\ )$ and $\boldsymbol{f_1} = \boldsymbol{Pois}(\ \lambda_1\ )$, then Equation 1 reduces to

$$C_t = \max\left(0, C_{t-1} + X_t - \frac{\lambda_1 - \lambda_0}{\ln(\lambda_1) - \ln(\lambda_0)}\right) \tag{3}$$

For the Poisson distribution the expected value equals the variance, $E(Y) = Var(Y) = \lambda$, so the Poisson-based CUSUM simultaneously monitors the mean and variance of the distribution.

C. COMPARING NORMAL- AND POISSON-BASED CUSUMS

No references were found in the literature to any studies comparing the performance of the normal- and Poisson-based CUSUM for the detection of rare diseases. Comparisons between Poisson-based CUSUM and other methods can be found in Han, Tsui, Ariyajunya, and Kim (2009); Rossi, Lampugnani, and Marchi (1999) compare the standardized Poisson CUSUM with functional transformation of the Poisson distribution

that produce a better normal approximation. Hawkins and Olwell (1998) assess the sensitivity of the Poisson CUSUM to deviations from the Poisson distribution.

When comparing Equations 2 and 3 it can be seen that normal- and Poisson – based CUSUM will perform equally when

$$\frac{\mu_1 - \mu_0}{2} \equiv \frac{\lambda_1 - \lambda_0}{2} = \frac{\lambda_1 - \lambda_0}{\ln(\lambda_1) - \ln(\lambda_0)}$$

Therefore, the normal- and Poisson-based CUSUMs are exactly the same when the denominators in the above equation are equal, which is the same as when

$$\frac{\lambda_1}{\lambda_0} = e^2 \approx 7.3891$$

Expressed in a more intuitive way, when the outbreak mean is approximately 7.40 times the nonoutbreak mean (i.e., when $\lambda_1 \approx 7.4 \times \lambda_0$) then the normal- and Poisson-based CUSUMs are exactly the same and thus, theoretically, will perform equally. This, of course, begs the question of what happens when the outbreak mean is either larger or smaller than 7.40 times the nonoutbreak mean.

1. Mean Relationship: Nonoutbreak and Outbreak Distributions

The mean-relationship between the nonoutbreak and outbreak distributions give us three particular cases to evaluate and compare the performance between normal- and Poisson-based CUSUM.

a. Case #1: $\lambda_1 = 7.40\ \lambda_0$

The outbreak mean is equal to 7.40 times the nonoutbreak mean. This case is a theoretical result that provides a reference point from which extreme case situations could be discriminated. It is worth noting that this is something of an extreme case in and of itself, in the sense that the mean occurrence for the outbreak is more than

seven times larger than the nonoutbreak mean, which is likely to be fairly obvious in and of itself. That is, an outbreak of this magnitude may be fairly obvious and not require the use of a statistical algorithm for detection.

b. Case #2: $\lambda_1 << 7.40\ \lambda_0$

The outbreak mean is much less than 7.40 times the nonoutbreak mean. For example, when the outbreak mean is almost equal to the nonoutbreak mean. This will be the case when the user desires to have an aggressive early detection policy, such as the biosurveillance system implementation in an urban area. It is also the case when an outbreak will not be obvious to medical and public health practitioners and likely require the use of a sensitive statistical algorithm to aid in detection.

c. Case #3: $\lambda_1 >> 7.40\ \lambda_0$

The outbreak mean is much greater than 7.40 times the nonoutbreak mean. This case is probably unlikely, but it would occur only when the user desires to detect outbreaks that are substantially larger than the nonoutbreak incidence rate. As with Case #1, this case is likely to be obvious and not require the use of a statistical algorithm for early event detection.

IV. COMPARISON METHODOLOGY

Bringing the material in Chapters I–III together, here we assess the performance of the normal- and Poisson-based CUSUMs for detecting Tularemia for Cases #1–#3 established at the end of Chapter III. The goal is to provide some perspective about how each algorithm performs relative to the other when trying to detect a rare disease and, in particular, to examine what the implications are of using a normal-based CUSUM when the data is Poisson distributed (as it is likely to be for rare diseases).

A. SCENARIO

As previously described, *F. tularensis* is a Category A biological agent due to its virulence, infectiousness, and ease of dissemination. As described in Chapter II, Tularemia occurs naturally, including in southern Monterey County. Thus, though it is rare, an occasional diagnosed case of Tularemia in and of itself is not sufficient evidence of either a large (naturally occurring) outbreak or of a bioterrorist attack.

1. Biological Agent

Disease: Tularemia

Cause Agent: *Francisella tularensis* type A

Incubation Period: 3 to 5 days (range 1 to 14 days)

Cases Occurrence: According to a *Poisson*(λ^*) distribution, a discrete distribution. The parameter λ^* is the mean number of cases actually observed in some population, which will be assumed constant for the purposes of this analysis. When $\lambda^* = \lambda_0$ then the observed mean is equal to the mean when there is no outbreak. However, when $\lambda^* > \lambda_0 \equiv \mu_0$ then there is an increased incidence in Tularemia and $\lambda^* \geq \lambda_1 \equiv \mu_1$ corresponds to an increase in the incidence that those who designed the CUSUM deemed important to detect quickly.

2. Biosurveillance Systems

Two biosurveillance systems are considered in this study.

a. System A

Uses the normal-based CUSUM (Equation 2) and finds the threshold h_A to achieve a desired ATFS assuming the disease occurrence distribution is approximately normal (e.g., for $X \sim Poisson(\lambda)$ the system operator (incorrectly) assumes $X \sim N(\lambda, \lambda)$).

This type of system is what would likely be used for detecting rare diseases outbreaks, incorrectly using the normal-based CUSUM to monitor data from a Poisson distribution.

b. System B

Uses the Poisson-based CUSUM (Equation 3) and finds the threshold h_B to a desired ATFS assuming the disease occurrence follows a Poisson distribution.

This system models the occurrence of outbreak of a rare disease more appropriately than System A because it does not assume any theoretical approximations, making it a good choice for comparisons with other systems.

B. METHODOLOGY

The idea is to gain some insight into what happens when a normal-based CUSUM (e.g., Equation 2) is used for detecting an outbreak of a rare disease in which the nonoutbreak observations can be assumed to follow a Poisson distribution. Under these conditions, the appropriate CUSUM to use would be Equation 3, but the CUSUM from Equation 2 is mistakenly applied.

The comparison metric used is the ATFS. This is the average time to false signal when there is no outbreak and it is also the average time to first signal when there is an outbreak. Under the assumption that the data is independently and identically distributed, these statistics are equivalent to the in-control and out-of-control average run length

metrics in statistical process control (Fricker, 2010a; Fricker, 2010b). When comparing the performance of the two methods, one first chooses thresholds so that both methods have equal average time to false signal, and then for various outbreak distributions, the method with the lower average time to first signal is considered better. Specifically, once the thresholds to achieve the desired ATFS are determined for each case, the average time to first signal for the normal- and Poisson-based CUSUMs are assessed across a variety of outbreaks means λ^* ($\lambda^* > \lambda_0$).

The effectiveness of Biosurveillance Systems A and B for the detection of a tularemia outbreak were simulated using an open source statistical software developed at Bell Laboratories, called R, available from http://www.r-project.org/. R was also used to determine the thresholds that achieved the desired ATFS and to simulate the Tularemia data used for comparison of the biosurveillance systems. The names of the R functions used are denoted in bold and blue color and the code is given in Appendix A.

The biosurveillance systems in use generally assume the data being surveilled is normally distributed and thus use normal-based CUSUM methods. However, rare diseases data do not conform to that assumption. Therefore, as Chpater V shows, the biosurveillance system will not behave as expected unless a different threshold is chosen. In order to compensate non-conformity of rare disease data to a normal-based CUSUM a process of calibration is used. A point of calibration P will be defined as the point where, when the mean rate of ocurrance (λ^*) is equal to nonoutbreak mean (λ_0), the thresholds of the two systems are set so that the have equal ATFSes.

The user of biosurveillance System A, which uses a normal-based CUSUM and who (incorrectly) assumes the data is normally distributed, sets a threshold (h_A) for the normal-based CUSUM under these assumptions to achieve a desired ATFS. The R code **fNormNorm2** is used to find this threshold.

The user of biosurveillance System B, the Poisson-based CUSUM and who (correctly) assumes the data is Poisson distributed, sets a threshold (h_B) for the Poisson-based CUSUM method to match the System A desired ATFS. The R code **fPoisPois2** is used to find this threshold.

At this point, Referred as stage 1, Biosurveillance Systems A and B are thought to be ready to be implemented to provide an early detection of a Tularemia outbreak. However, the Tularemia disease incidence rate in the United States is less than 150 cases per year (Weinberg & Branda, 2010; CDC, 2002), which gives an indication that the data cannot be assumed to be normally distributed and therefore System A may not perform as expected.

Stage 2 begins once Biosurveillance Systems A and B are implemented in the field to provide an early detection of a Tularemia outbreak resulting from either natural causes or bioterrorist attack. Because Biosurveillance System A may not behave as expected, and this could be seen by comparing the achieved ATFS to the desired ATFS at the calibration point P, the user will likely have to adjust the threshold to achieve the desired ATFS performance.

That is, the user of System A, noticing the difference between achieved ATFS and the desired ATFS, re-calibrates the system (i.e., changes the threshold); this is referred as Stage 3. The threshold adjustment to achieve the desired ATFS is done using the **fNormPois2** R code. The new settings for the normal-based CUSUM method is called Biosurveillance System A' with threshold $h_{A'}$.

The results are depicted in a series of two graphs for each mean relationship case, where the first graph for each shows Systems A and B being implemented in an environment with varying occurrence outbreak mean; the second graph, for each case, shows the improvement or lack of it of System A' in comparison to Systems A and B.

C. ESTIMATING LAMBDA

For an implementation of the system the nonoutbreak (λ_0) and outbreak means (λ_1) need to be estimated. The nonoutbreak mean λ_0 can be estimated from historical data and the outbreak mean λ_1 will be set depending on the level of biosurveillance aggressiveness desired according to the biosurveillance established policy.

Studying the incidence rate within the United States, it can be seen that 125 tularemia cases per year were recorded in the 2000–2008 period (Weinberg & Branda,

2010). The biosurveillance system is thought to be operating in a daily cycle,which is the data fed into the CUSUM algorithm is on a daily basis, the value of lambda should be computed per day.

One initial, simple and intuitive approach is to consider only the number of cases within the entire United States per year. Using the properties of the Poisson distribution, the number of cases reported per year are translated to a number of cases per day that can be considered the nonoutbreak mean $\lambda_{0,}$ and the outbreak mean will be estimated based on the aggressiveness of the surveillance policy. In a situation where a system is being designed for a specific county or city, the population considered will be the local population being monitored and the local incidence rate.

To find a rough and broad estimate for the nonoutbreak mean λ_0, the number of yearly cases is divided by the number of days in a year (disregarding leap years) to get the mean of the number of cases per day. Since the mean of the Poisson distribution is lambda, the procedure gives a simple estimation for the parameter used in the CUSUM algorithm. In the Tularemia scenario, the nonoutbreak mean is 0.3425 and, therefore, the number of days with non zero cases in a year are very few. This is the first indication that the data is better adjusted to a Poisson distribution than a normal distribution, and that transforming the data to approximate normality may not be appropriate for rare diseases surveillance.

THIS PAGE INTENTIONALLY LEFT BLANK

V. RESULTS AND ANALYSIS

Since the user sets the desired-ATFS based on desired system performance, since the incidence rate for nonoutbreak situations gives an estimate of λ_0 and the biosurveillance system policy specifies λ_1 and since the theory shows that the importance relies in the means relationship rather than the means values, in the simulations the means can be chosen arbitrarily.

To keep the computations reasonable, both in terms of the simulation run time and the use of computer memory, the values of nonoutbreak means are fixed at 0.100 (λ_0 = 0.100). The value of the outbreak mean λ_1, will depend on the means relationship of the case in consideration. The number of loops chosen to run the simulations was set so that the standard error for the target-ATFS is less than 0.100.

Given that the Tularemia data is discrete, one cannot assume that resulting ATFS from the simulation will be continuous. Preliminary runs were conducted to select the appropriate target ATFS as close to the desired-ATFS (maintain equal to 80.00) to allow comparisons between the three considered biosurveillance systems without unnecessary efforts. Details of these preliminary runs are not given here; however, they can be provided upon request to the author. A sample threshold search is shown in appendix C, where it can bee seen discontinuity in the ATFS suggesting that selection of an ATFS may not be arbitrary and more research is needed to understand completely the situation.

From the simulation results, Appendix B, when comparing systems effectiveness at $\lambda^* = \lambda_0$, the calibration point *P*, the reference ATFS is the target-ATFS of System B that is a consequence from computing the threshold for a desired ATFS of that system at each case of the mean relationship. However, for $\lambda^* > \lambda_0$, the reference ATFS should be the achieved-ATFS of System B at that occurrence outbreak mean (λ^*) because that ATFS and the achieved-ATFS of the other systems, are computed using the same set of parameters. See Table 3 for a simplified process explanation to assess the effectiveness between normal- and Poisson-based CUSUM.

Stage		ATFS	Threshold
Stage 1	Based on the aggressiveness of surveillance policy select desired-ATFS to find the appropriate threshold h that will give an ATFS, target-ATFS, close to the desired ATFS.	desired-ATFS target-$ATFS_{A, B}$	$h_{A, B}$
Stage 2	For the first simulation run target-$ATFS_B$ is the reference ATFS at the calibration point P where $\lambda^* = \lambda_0$. Away from the calibration point P, that is for $\lambda^* > \lambda_0$ the reference is the achieved-$ATFS_B$ at λ^*	target-$ATFS_{A,B}$, at point P achieved-$ATFS_{A,B}$	$h_{A, B}$
Stage 3	Adjustment of System A threshold h_A. System A' is System A with adjusted threshold.	target-$ATFS_{A,B,A'}$, at point P achieved-$ATFS_{A,B,A'}$	$h_{A, B, A'}$

Table 3. Stage Process to Assess Effectiveness of Normal- and Poisson-Based CUSUM

A. CASE #1

In this case the mean relationship is $\lambda_1 = 7.40\ \lambda_0$, where the theoretical behavior of the normal- and Poisson-based CUSUM coincide.

1. Case #1: Stage 1

With outbreak mean of $\lambda_1 = 0.740$, the resulting target-ATFS for each biosurveillance systems A and B are 78.90 and 78.75, respectively, with threshold values of 0.64 and 1.36, respectively. The computational results are shown in Table 4.

System A			System B		
Target ATFS	Standard Error	Threshold $\boldsymbol{h_A}$	Target ATFS	Standard Error	Threshold $\boldsymbol{h_B}$
78.9001	0.0773	0.6378	78.7518	0.0777	1.3604

Table 4. Case #1: Stage 1. Thresholds Implementation Parameters

2. Case #1: Stage 2

Once the systems are fielded or deployed to provide an early detection of Tularemia, Figure 1 shows that only System B achieves an ATFS near to the target-$ATFS_B$ at the calibration point P.

In particular, in Figure 2 notes that System B achieves an ATFS of 78.64 while System A achieves an ATFS of 10.51—well below that of System B and what is desired. In System A, the difference between target-ATFS of System B and the achieved-ATFS of System A at the calibration point *P* is 68.24. Thus, System A will have a false positive signal rate on the average almost eight times what is desired.

Now, Figure 2 also shows that System A will detect outbreaks more quickly than System B. This follows because the ATFS for System A is much lower than for System B for all $\lambda^* > \lambda_0$. However, this could simply be an artifact of the fact that the threshold for System A is set inappropriately low.

3. Case #1: Stage 3

In this stage, the threshold $h_{A'}$ is determined to be 1.08, achieving an ATFS of 78.26 that is near to the target-ATFS of System B at the calibration point *P*. Table 5 shows the parameters of the three biosurveillance systems in consideration; in Figure 3, it is clear that System A' and System B perform almost exactly the same.

In fact, this result should not be surprising since the CUSUMs from Equations 2 and 3 are exactly the same under these conditions. However, it is important to note that System A will give an excessive number of false positive signals if there is an assumption the data were normally distributed and the threshold is set incorrectly.

System A			System B			System A'		
Target ATFS	Standard Error	Threshold $\boldsymbol{h_A}$	Target ATFS	Standard Error	Threshold $\boldsymbol{h_B}$	Target ATFS	Standard Error	Threshold $\boldsymbol{h_{A'}}$
78.9001	0.0773	0.6378	78.7518	0.0777	1.3604	78.2039	0.0769	1.0805

Table 5. Case #1: Stage 3. Thresholds Implementation Parameters

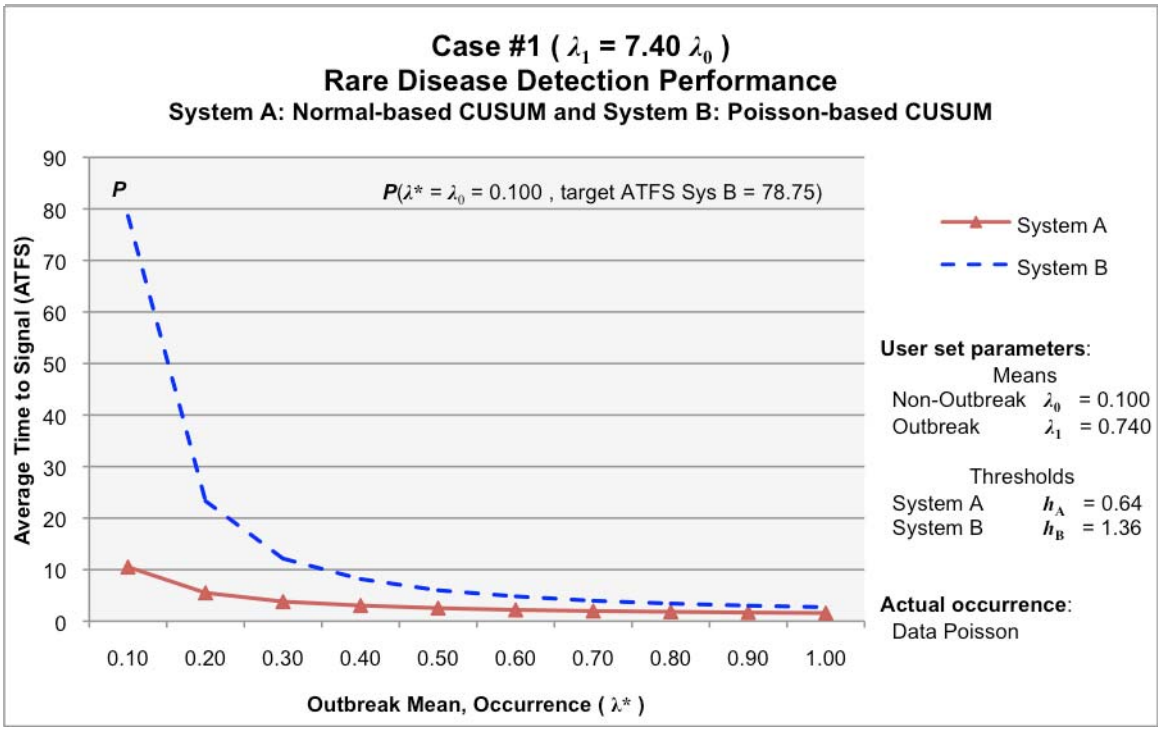

Figure 2. Case #1 Stage 2

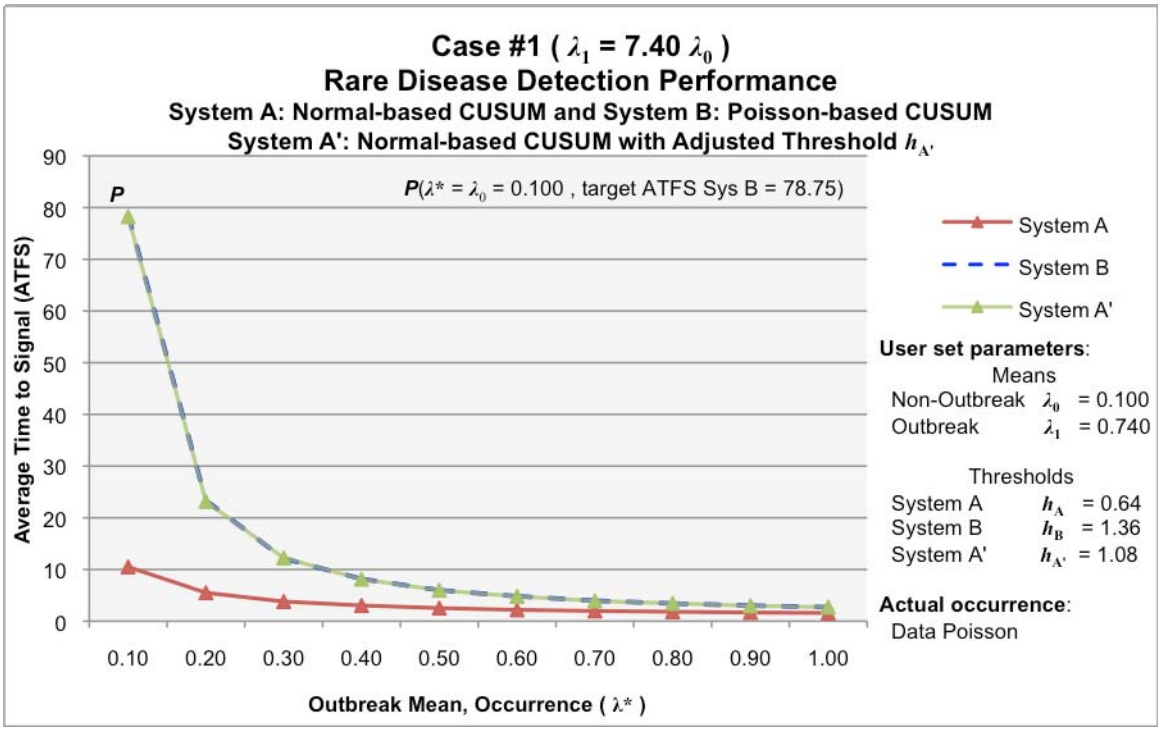

Figure 3. Case #1 Stage 3

B. CASE #2

In this case, the mean relationship is $\lambda_1 << 7.40\ \lambda_0$, where an aggressive early detection policy is desired for the biosurveillance system. This case could be applicable when there is knowledge of a probable bioterrorist attack (intentional release of biological agent) or in other situations where the user wants a system sensitive to detecting a Tularemia outbreak.

1. Case #2: Stage 1

With outbreak mean of $\lambda_1 = 0.105$, very close to λ_0, the resulting target-ATFS for each biosurveillance systems A and B are 81.58 and 81.08 respectively with threshold values of 8.11 and 2.39, respectively. The computational results from finding the thresholds to be used are shown in Table 6.

System A			System B		
Target ATFS	Standard Error	Threshold h_A	Target ATFS	Standard Error	Threshold h_B
81.5763	0.0877	8.1100	81.0823	0.0730	2.3851

Table 6. Case #2: Stage 1. Threshold Implementation Parameters

2. Case #2: Stage 2

Once the systems are fielded or deployed to provide an early detection of tularemia, Figure 3 shows that the System A achieved-ATFS is greater than the target-ATFS of System B at the calibration point *P*. This result is the opposite of that obtained in Case #1, where System A was signaling below the target-$ATFS_B$.

System B signals at the achieved-ATFS of 81.45, while System A signals at an achieved-ATFS of 90.19. In System A, the difference between the target-$ATFS_B$ of System B and achieved-ATFS of System A at the calibration point *P* is 9.11. This difference of 9.11 represents about a 10% increase in the time between false positive signals. Furthermore, as shown in Figure 4, System A has a larger ATFS than System B for all $\lambda^* > \lambda_0$. Thus, in this configuration, System A will be slower than System B at

detecting an outbreak. For Tularemia with an incubation time of 3 to 5 days, such delays in detection would increase the time for post-exposure prophylaxis that will result in a greater number of persons without having the correct antibiotic therapy. This delay will more likely increase the number of fatalities within the exposed population and the level of resources used to avoid further casualties. However, this delay could simply be an artifact of the fact that the threshold for System A is set incorrectly.

3. Case #2: Stage 3

In this stage, the threshold $h_{A'}$ is determined to be 7.73, achieving an ATFS of 81.68 that is near to the target-$ATFS_B$ of System B. Table 7 shows the parameters of the three considered biosurveillance systems.

System A			System B			System A'		
Target ATFS	Standard Error	Threshold $\boldsymbol{h_A}$	Target ATFS	Standard Error	Threshold $\boldsymbol{h_B}$	Target ATFS	Standard Error	Threshold $\boldsymbol{h_{A'}}$
81.5763	0.0877	8.1100	81.0823	0.0730	2.3851	81.6840	0.0967	7.7319

Table 7. Case #2: Stage 3. Thresholds Implementation Parameters

As shown in Figure 5, System A', the normal-based CUSUM with threshold $h_{A'}$ of 7.73, achieves an ATFS of 81.73; now System A' is thought to be calibrated at point *P*. However, with an increase of the occurrence outbreak mean, System A' continues to show a greater ATFS values than System B for all $\lambda^* > \lambda_0$. The result is that, even after correcting the biosurveillance system, as in System A', the CUSUM incorrectly assuming the data is normally distributed is significantly slower to detect a Tularemia outbreak by something on the order of 5 to15 days' delay in detection.

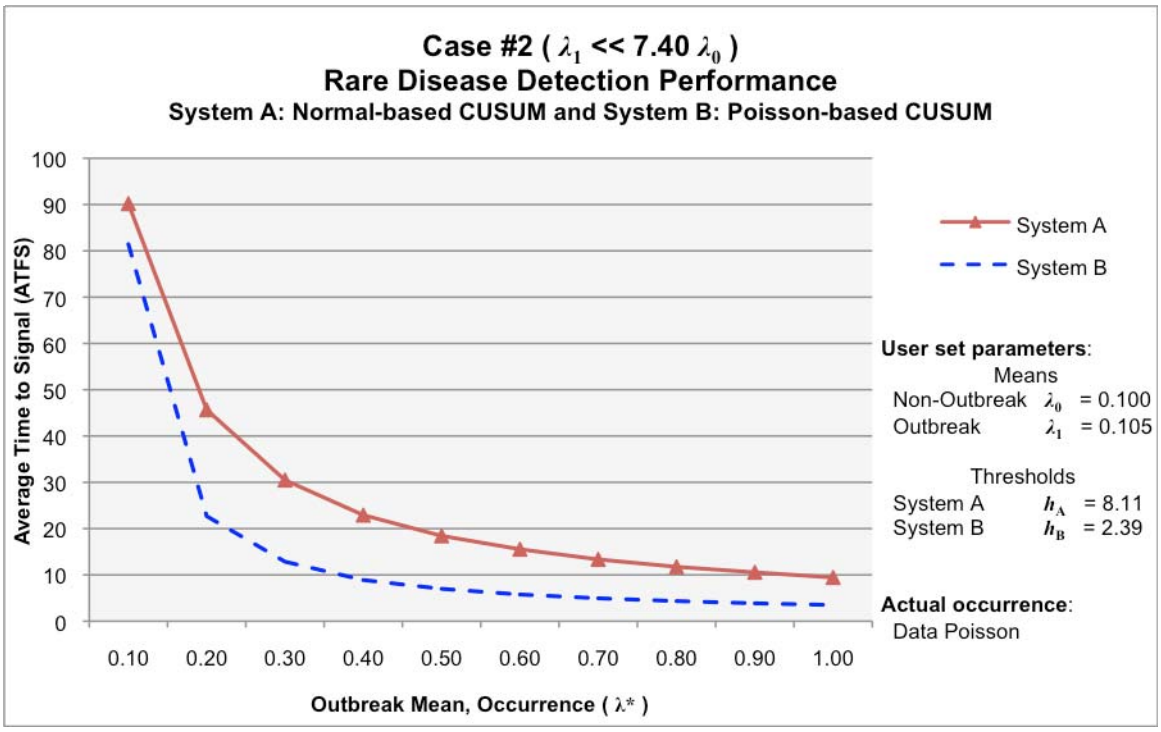

Figure 4. Case #2 Stage 2

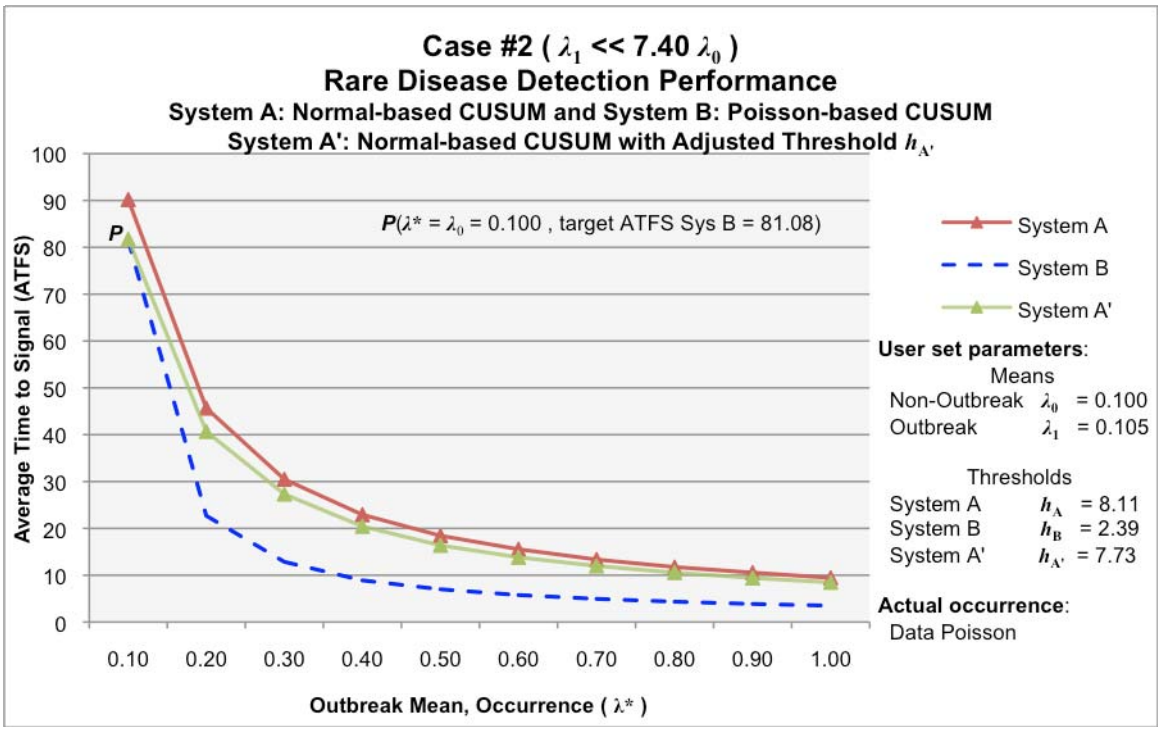

Figure 5. Case #2 Stage 3

C. CASE #3

In this case, the mean relationship is $\lambda_1 >> 7.40\ \lambda_0$, where a less-aggressive early detection policy is desired for the biosurveillance system. This case may be applicable, for example, for a biosurveillance system implementation in a rural area, where there is knowledge of natural occurrence cases. This happens in those states where tularemia naturally occurs, such as Missouri, Arkansas, Oklahoma, and Massachusetts. A particular area in Massachusetts is Martha's Vineyard, where the majority of the cases in the United States have occurred (Weinberg & Branda, 2010).

1. Case #3: Stage 1

With outbreak mean of $\lambda_1 = 1.480$ (twice the mean relationship of Case #1), the resulting target-ATFS for each of Biosurveillance Systems A and B are 82.62 and 82.08 respectively with threshold values of 0.12 and 0.49, respectively. The computational results for determining the thresholds are shown in Table 8.

System A			System B		
Target ATFS	Standard Error	Threshold h_A	Target ATFS	Standard Error	Threshold h_B
82.6174	0.0822	0.1238	82.0748	0.0809	0.4879

Table 8. Case #3: Stage 1. Thresholds Implementation Parameters

2. Case #3: Stage 2

Once the systems are fielded or deployed to provide an early detection of tularemia, Figure 5 shows that only System B achieves the target-$ATFS_B$ at the calibration point *P*.

That is, as shown in Figure 6, System B achieves an ATFS of 81.17, while System A achieves an ATFS of 10.71. In System A, the difference between target-$ATFS_B$ of System B and achieved-$ATFS_A$ of System A at the calibration point *P* is 71.37 This result is similar to that seen in Case #1.

For a disease like Tularemia that occurs naturally in the rural areas of the Midwestern of United States and Martha's Vineyard in Massachusetts, the biosurveillance policy might be to set a high outbreak mean to account for those naturally occurring cases. However, if System A starts to behave as shown in Figure 6, the user will lose trust in the effectiveness of the biosurveillance system and, perhaps, begin to ignore the system's signals.

3. Case #3: Stage 3

In this stage, the threshold $h_{A'}$ is determined to be 0.31 achieving an ATFS of 81.48, which is very close to the target-$ATFS_B$. Table 9 shows the parameters of the three biosurveillance systems in consideration, where it is clear that System A' and System B perform almost exactly the same. See Figure 7.

System A			System B			System A'		
Target ATFS	Standard Error	Threshold $\boldsymbol{h_A}$	Target ATFS	Standard Error	Threshold $\boldsymbol{h_B}$	Target ATFS	Standard Error	Threshold $\boldsymbol{h_{A'}}$
82.6174	0.0822	0.1238	82.0748	0.0809	0.4879	82.2159	0.0812	0.3124

Table 9. Case #3: Stage 3. Thresholds Implementation Parameters

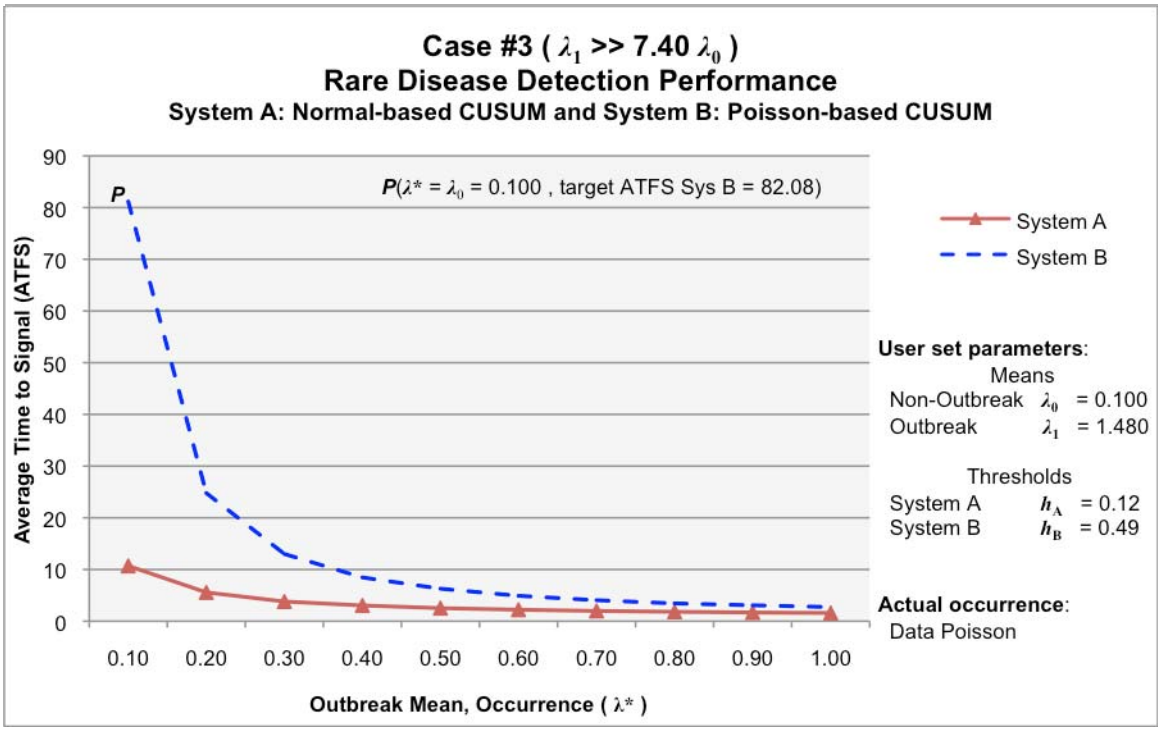

Figure 6. Case #3 Stage 2

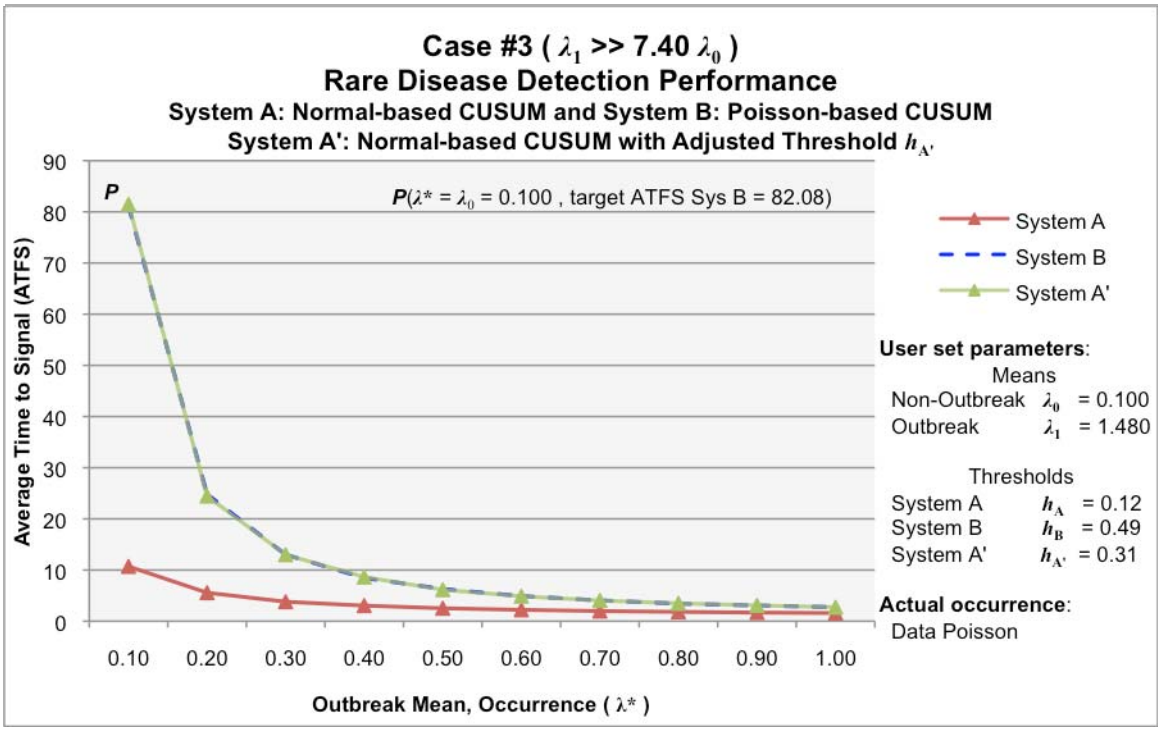

Figure 7. Case #3 Stage 3

VI. CONCLUSIONS

A. CONCLUSION #1

When the rate of disease, λ_0, is very low and the occurrence counts are Poisson distributed, $X \sim Pois(\lambda_0)$, and the outbreak mean, λ_1, manifests as only a small increase in the rate of disease, $\lambda_1 << 7.40\ \lambda_0$., then the normal-based CUSUM can be significantly slower to signal an outbreak (on average) compared to the equivalent Poisson-based CUSUM.

Thus, the incorrect use of normal-based CUSUM can result in unacceptable delay in detection. To monitor a rare disease, such as Tularemia, a Poisson-based CUSUM should be included in biosurveillance systems. The examples herein show potential delays in signaling on the order of weeks when normal-based CUSUMs are used to monitor Poisson data.

B. CONCLUSION #2

When, the rate of disease, λ_0, is very low, the occurrence counts are Poisson distributed, $X \sim Pois(\lambda_0)$, and the outbreak, λ_1, is significantly larger than rate of disease, $\lambda_1 \geq 7.40\ \lambda_0$, then, the normal-based CUSUM performs as well as the Poisson-based CUSUM if the threshold h, is set appropriately to achieve equivalent ATFS performance at $\lambda^* = \lambda_0$. However, if the threshold h is set incorrectly then normal-based CUSUM has an excessively high false signal rate. This echoes a real problem with existing biosurveillance systems.

C. CONCLUSION #3

This work provides an objective methodology to determine the thresholds for a normal- and Poisson-based CUSUM being used for a biosurveillance system intended to detect emerging infectious disease and any disease caused by a Category A biological agents.

D. FUTURE RESEARCH OPPORTUNITIES

More research should be devoted to the study of discontinuity in the average time to signal (ATFS) for varying threshold (h) when applying normal- and Poisson-based CUSUM with Poisson distributed data, as it was observed during the search for an appropriate threshold, suggesting that the selection of the ATFS may not be arbitrary. See Appendix C for an example of discontinuities observed in the threshold search run results.

APPENDIX A. R CODE

A. AVERAGE TIME TO SIGNAL - CUSUM METHOD

```
# function nomenclature
#  f  to  denote  function;  distribution-based  CUSUM;  distribution  data
(observations)
# ComplexLevel: {1:vector, 2:matrix}
# fPoisPois1        Poisson-based CUSUM with Poisson observations
# fNormNorm1        normal-based CUSUM with normal observations
# fNormPois1        normal-based CUSUM with Poisson observations
# fPoisPois2        Find threshold for a desire.ATFS
# fNormNorm2        Find threshold for a desire.ATFS
# fNormPois2        Find threshold for a desire.ATFS
# fIn.Field         Simulation for Systems A, B and A' for varying occurrence
outbreak mean (lambda.star) with plot

# Poisson-based CUSUM with Poisson observations
# Parameters:
# h                 Threshold
# nr.loops          number of loops
# lambda0           nonoutbreak mean
# lambda1           outbreak mean
# lambda.star       occurrence outbreak mean (in-field)
# Return:           ATFS and standard error in a data frame structure

fPoisPois1 <- function(h, nr.loops, lambda0, lambda1, lambda.star=lambda0) {
   delay <- rep(0, nr.loops);
   k <- (lambda1 - lambda0)/(log(lambda1) - log(lambda0))

   for(i in 1:nr.loops) {
      Ct <- 0;
      counter <- 0;

      while(Ct < h) {
         obs <- rpois(1,lambda.star)
         Ct <- max(0, Ct + obs - k)
         counter <- counter + 1
         }

      delay[i] <- counter
      }

   atfs.PoisPois <- mean(delay)
   se.PoisPois <- sd(delay)/sqrt(nr.loops)

   antwort.data <- data.frame(atfs.PoisPois, se.PoisPois)
   return(antwort.data)
   }

# normal-based CUSUM with normal observations
# Parameters:
# h                       Threshold
# nr.loops          number of loops
# lambda0           nonoutbreak mean
# lambda1           outbreak mean
# lambda.star       occurrence outbreak mean (in-field)
```

```
# Return:           ATFS and standard error in a data frame structure

fNormNorm1 <- function(h, nr.loops, lambda0, lambda1, lambda.star=lambda0) {
   delay <- rep(0, nr.loops);
   k <- (lambda1-lambda0)/2

   for(i in 1:nr.loops) {
      Ct <- 0;
      counter <- 0;

      while(Ct < h) {
         obs <- rnorm(1, lambda.star, sqrt(lambda.star))
         Ct <- max(0, Ct + obs - k)
         counter <- counter + 1
         }

      delay[i] <- counter
      }

   atfs.NormNorm <- mean(delay)
   se.NormNorm <- sd(delay)/sqrt(nr.loops)

   antwort.data <- data.frame(atfs.NormNorm, se.NormNorm)
   return(antwort.data)
   }

# normal-based CUSUM with Poisson observations
# Parameters:
# h                      Threshold
# nr.loops          number of loops
# lambda0           nonoutbreak mean
# lambda1           outbreak mean
# lambda.star       occurrence outbreak mean (in-field)
# Return:           ATFS and standard error in a data frame structure

fNormPois1 <- function(h, nr.loops, lambda0, lambda1, lambda.star=lambda0) {
   delay <- rep(0, nr.loops);
   k <- (lambda1-lambda0)/2

   for(i in 1:nr.loops) {
      Ct <- 0;
      counter <- 0;

      while(Ct < h) {
         obs <- rpois(1, lambda.star)
         Ct <- max(0, Ct + obs - k)
         counter <- counter + 1
         }

      delay[i] <- counter
      }

   atfs.NormPois <- mean(delay)
   se.NormPois <- sd(delay)/sqrt(nr.loops)

   antwort.data <- data.frame(atfs.NormPois, se.NormPois)
   return(antwort.data)
   }
```

B. THRESHOLD SEARCH

```
# Find threshold for a desire.ATFS
# Parameters:
# lb.h              lower bound  Threshold
# ub.h              upper bound Threshold
# step.h            step size for Threshold search
# nr.loops          number of loops
# lambda0           nonoutbreak mean
# lambda1           outbreak mean
# lambda.star       occurrence outbreak mean (in-field)
# desire.atfs       desire ATFS
# epsilon           exit criteria
# Return:           Threshold, target ATFS, target standard error in a matrix
structure
# Function Calls
#                   fPoisPois1
#                   fPoisPois2

fPoisPois2 <- function(lb.h, ub.h, step.h=0.1, nr.loops=100, lambda0, lambda1,
lambda.star=lambda0, desire.atfs=80, epsilon=0.01) {
   delay <- rep(0, nr.loops);
   threshold <- seq(lb.h, ub.h, step.h)
   antwort.matrix <- matrix(nrow = length(threshold), ncol = 3)

   index <- 1
   old.h <- 0
   new.h <- lb.h
   for(i in threshold) {
      old.h <- new.h
      new.h <- i

      antwort.data <- fPoisPois1(i, nr.loops, lambda0, lambda1, lambda.star)

      antwort.matrix[index,1] <- i
      antwort.matrix[index,2] <- antwort.data[1,1]
      antwort.matrix[index,3] <- antwort.data[1,2]

      print(antwort.matrix[index,])

      if(antwort.data[1,1] > desire.atfs) {
         dimnames(antwort.matrix)   <-   list(threshold   ,   c("h.PoisPois",
         "atfs.PoisPois", "se.PoisPois")); print(antwort.matrix)

         if(step.h > 0.0001) {
            step.h <- step.h/10
            }

         nr.loops <- nr.loops*10
         subtitle <- c(step.h, nr.loops)
         print("step.h nr.loops"); print(subtitle)

         fPoisPois2(old.h, new.h + step.h, step.h, nr.loops, lambda0, lambda1,
         lambda.star, desire.atfs)
         }

      index <- index + 1
      }

   if(abs(antwort.data[1,1] - desire.atfs) > epsilon) {
```

```
        dimnames(antwort.matrix)    <-    list(threshold    ,    c("h.PoisPois",
        "atfs.PoisPois", "se.PoisPois")); print(antwort.matrix)

        fPoisPois2(old.h - step.h, new.h + 10*step.h, step.h, nr.loops, lambda0,
        lambda1, lambda.star, desire.atfs)
        }

    dimnames(antwort.matrix)     <-     list(threshold     ,     c("h.PoisPois",
    "atfs.PoisPois", "se.PoisPois"))

    return(antwort.matrix)
    }

# Find threshold for a desire.ATFS
# Parameters:
# lb.h              lower bound  Threshold
# ub.h              upper bound Threshold
# step.h            step size for Threshold search
# nr.loops          number of loops
# lambda0           nonoutbreak mean
# lambda1           outbreak mean
# lambda.star       occurrence outbreak mean (in-field)
# desire.atfs       desire ATFS
# epsilon           exit criteria
# Return:           Threshold, target ATFS, target standard error in a matrix
structure
# Function Calls
#                   fNormNorm1
#                   fNormNorm2

fNormNorm2 <- function(lb.h, ub.h, step.h=0.1, nr.loops=100, lambda0, lambda1,
lambda.star=lambda0, desire.atfs=80, epsilon=0.01) {
    delay <- rep(0, nr.loops);
    threshold <- seq(lb.h, ub.h, step.h)
    antwort.matrix <- matrix(nrow = length(threshold), ncol = 3)

    index <- 1
    old.h <- 0
    new.h <- lb.h
    for(i in threshold) {
        old.h <- new.h
        new.h <- i
        antwort.data <- fNormNorm1(i, nr.loops, lambda0, lambda1, lambda.star)

        antwort.matrix[index,1] <- i
        antwort.matrix[index,2] <- antwort.data[1,1]
        antwort.matrix[index,3] <- antwort.data[1,2]

        print(antwort.matrix[index,])

        if(antwort.data[1,1] > desire.atfs) {
            dimnames(antwort.matrix)    <-    list(threshold,    c("h.NormNorm",
            "atfs.NormNorm", "se.NormNorm")); print(antwort.matrix)

            if(step.h > 0.0001) {
                step.h <- step.h/10
                }

            nr.loops <- nr.loops*10
            subtitle <- c(step.h, nr.loops)
            print("step.h nr.loops"); print(subtitle)
```

```
            fNormNorm2(old.h, new.h + step.h, step.h, nr.loops, lambda0, lambda1,
            lambda.star, desire.atfs)
            }

        index <- index + 1
        }

    if(abs(antwort.data[1,1] - desire.atfs) > epsilon) {
        dimnames(antwort.matrix)      <-      list(threshold,      c("h.NormNorm",
        "atfs.NormNorm", "se.NormNorm")); print(antwort.matrix)

        fNormNorm2(old.h - step.h, new.h + 10*step.h, step.h, nr.loops, lambda0,
        lambda1, lambda.star, target.atfs)
        }

    dimnames(antwort.matrix) <- list(threshold, c("h.NormNorm", "atfs.NormNorm",
    "se.NormNorm"))

    return(antwort.matrix)
    }


# Find threshold for a desire.ATFS
# Parameters:
# lb.h              lower bound   Threshold
# ub.h              upper bound Threshold
# step.h            step size for Threshold search
# nr.loops          number of loops
# lambda0           nonoutbreak mean
# lambda1           outbreak mean
# lambda.star       occurrence outbreak mean (in-field)
# desire.atfs       desire ATFS
# epsilon           exit criteria
# Return:           Threshold, target ATFS, target standard error in a matrix
structure
# Function Calls
#                         fNormPois1
#                         fNormPois2

fNormPois2 <- function(lb.h, ub.h, step.h=0.1, nr.loops=100, lambda0, lambda1,
lambda.star=lambda0, desire.atfs=60, epsilon=0.01) {
    delay <- rep(0, nr.loops);
    threshold <- seq(lb.h, ub.h, step.h)
    antwort.matrix <- matrix(nrow = length(threshold), ncol = 3)

    index <- 1
    old.h <- 0
    new.h <- lb.h
    for(i in threshold) {
        old.h <- new.h
        new.h <- i
        antwort.data <- fNormPois1(i, nr.loops, lambda0, lambda1, lambda.star)

        antwort.matrix[index,1] <- i
        antwort.matrix[index,2] <- antwort.data[1,1]
        antwort.matrix[index,3] <- antwort.data[1,2]

        print(antwort.matrix[index,])

        if(antwort.data[1,1] > desire.atfs) {
            dimnames(antwort.matrix)    <-    list(threshold    ,    c("h.NormPois",
            "atfs.NormPois", "se.NormPois")); print(antwort.matrix)
```

```
        if(step.h > 0.0001) {
           step.h <- step.h/10
           }

        nr.loops <- nr.loops*10
        subtitle <- c(step.h, nr.loops)
        print("step.h nr.loops"); print(subtitle)

        fNormPois2(old.h - step.h, new.h + step.h, step.h, nr.loops, lambda0,
        lambda1, lambda.star, desire.atfs)
        }

     index <- index + 1
     }

  if(abs(antwort.data[1,1] - desire.atfs) > epsilon) {
     dimnames(antwort.matrix)    <-    list(threshold    ,    c("h.NormPois",
     "atfs.NormPois", "se.NormPois")); print(antwort.matrix)

     fNormPois2(old.h, new.h + 10*step.h, step.h, nr.loops, lambda0, lambda1,
     lambda.star, desire.atfs)
     }

  dimnames(antwort.matrix)     <-     list(threshold     ,     c("h.NormPois",
  "atfs.NormPois", "se.NormPois"))

  return(antwort.matrix)
  }
```

C. SIMULATION VARYING OCCURRENCE OUTBREAK MEAN

```
# Simulation for Systems A, B and A' for varying occurrence outbreak mean
(lambda.star) with plot
# Parameters:
# h1                Threshold System A
# h2                Threshold System B
# h3                Threshold System A'
# nr.loops          number of loops
# lambda0           nonoutbreak mean
# lambda1           outbreak mean
# star.max          maximum occurrence outbreak mean lambda.star
# Return:           ATFS for each system and difference system B - other system
in a matrix structure
# Function Calls
#                   fPoisPois2
#                   fNormNorm2
#                   fNormPois2

fIn.Field <- function(h1, h2, h3, nr.loops, lambda0, lambda1, star.max=3.0) {
   lambda.star <- seq(0.1, star.max, 0.1)
   atfs.h1 <- rep(0,length(lambda.star))
   atfs.h2 <- rep(0,length(lambda.star))
   atfs.h3 <- rep(0,length(lambda.star))

   for(i in 1:length(lambda.star)) {
      # System A
      antwort.h1 <- fNormPois1(h1, nr.loops, lambda0, lambda1, lambda.star[i])
      atfs.h1[i] <- antwort.h1[1,1]
```

```
    # System B
    antwort.h2 <- fPoisPois1(h2, nr.loops, lambda0, lambda1, lambda.star[i])
    atfs.h2[i] <- antwort.h2[1,1]

    # System A'
    antwort.h3 <- fNormPois1(h3, nr.loops, lambda0, lambda1, lambda.star[i])
    atfs.h3[i] <- antwort.h3[1,1]
    }

plot(lambda.star, atfs.h2, type = "l", lty = 4, col = 4, xlab =
"lambda^star", ylab = "Avg Time to Signal", ylim = c(0,100))

lines(lambda.star, atfs.h1, lty = 1, col = 2)
lines(lambda.star, atfs.h3, lty = 1, col = 3)

points(lambda.star, atfs.h2, pch = 1)
points(lambda.star, atfs.h1, pch = 4)
points(lambda.star, atfs.h3, pch = 5)

diff.atfs.h2h1 <- atfs.h2-atfs.h1
diff.atfs.h2h3 <- atfs.h2-atfs.h3

antwort.data <- data.frame(atfs.h1, atfs.h2, atfs.h3, diff.atfs.h2h1,
diff.atfs.h2h3)

return(antwort.data)
}
```

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. SIMULATION TABLE RESULTS

A. CASE #1

Occurrence Outbreak Mean (λ^*)	achieved-ATFS			achieved-$ATFS_B$ - achieved-$ATFS_{A,A'}$		target-$ATFS_B$ Sys B: 78.7518	
						target-$ATFS_B$ – achieved-$ATFS_{A,A'}$	
	System A	System B	System A'	System A	System A'	System A	System A'
0.10	10.5088	78.6372	78.2559	68.1284	0.3813	68.2430	0.4959
0.20	5.5022	23.3129	23.1702	17.8107	0.1427		
0.30	3.8064	12.1435	12.2490	8.3371	-0.1055		
0.40	3.0441	8.1768	8.1379	5.1327	0.0389		
0.50	2.5424	5.9954	5.9770	3.4530	0.0184		
0.60	2.1923	4.8213	4.8075	2.6290	0.0138		
0.70	1.9808	3.9615	3.9192	1.9807	0.0423		
0.80	1.8281	3.4237	3.4525	1.5956	-0.0288		
0.90	1.6836	3.0184	2.9805	1.3348	0.0379		
1.00	1.5865	2.7210	2.7317	1.1345	-0.0107		

B. CASE #2

Occurrence Outbreak Mean (λ^*)	achieved-ATFS			achieved-ATFS$_B$ - achieved-ATFS$_{A,A'}$		target-ATFS Sys B: 81.0823	
	System A	System B	System A'	System A	System A'	target-ATFS$_B$ – achieved-ATFS$_{A,A'}$ System A	System A'
0.10	90.1933	81.4491	81.7306	-8.74	-0.28	-9.11	-0.65
0.20	45.6744	22.6916	40.6759	-22.98	-17.98		
0.30	30.4880	12.8553	27.3096	-17.63	-14.45		
0.40	22.9179	8.8971	20.4274	-14.02	-11.53		
0.50	18.4341	6.9965	16.3710	-11.44	-9.37		
0.60	15.5456	5.7662	13.8204	-9.78	-8.05		
0.70	13.3474	4.9518	11.9883	-8.40	-7.04		
0.80	11.7519	4.3496	10.5585	-7.40	-6.21		
0.90	10.5558	3.8651	9.4035	-6.69	-5.54		
1.00	9.4429	3.5048	8.4776	-5.94	-4.97		

C. CASE #3

Occurrence Outbreak Mean (λ^*)	achieved-ATFS			achieved-ATFS$_B$ - achieved-ATFS$_{A,A'}$		target-ATFS Sys B: 82.0748	
	System A	System B	System A'	System A	System A'	target-ATFS$_B$ – achieved-ATFS$_{A,A'}$ System A	System A'
0.10	10.7054	81.1671	81.4763	70.46	-0.31	71.37	0.60
0.20	5.5690	24.7847	24.4483	19.22	0.34		
0.30	3.8087	13.0049	12.9909	9.20	0.01		
0.40	3.0462	8.4658	8.6337	5.42	-0.17		
0.50	2.5315	6.2885	6.1618	3.76	0.13		
0.60	2.2192	4.9118	4.8847	2.69	0.03		
0.70	1.9817	4.0602	4.0452	2.08	0.02		
0.80	1.8255	3.4414	3.4975	1.62	-0.06		
0.90	1.6860	3.0802	3.0972	1.39	-0.02		
1.00	1.5913	2.7360	2.7438	1.14	-0.01		

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C. THRESHOLD SEARCH EXAMPLE

Threshold search for normal-based CUSUM and data Poisson distributed

```
> fNormPois12(0.00001, 10, 0.1, 100, 0.1, 0.105, desired.atfs=80)
        h.NormPois atfs.NormPois se.NormPois
1e-05      0.00001         10.90   0.9909184
0.10001    0.10001         10.44   0.9708187
0.20001    0.20001          9.50   0.7533762
0.30001    0.30001         10.43   1.0186399
0.40001    0.40001          9.55   1.0321474
0.50001    0.50001          9.22   0.8001995
0.60001    0.60001          9.36   0.8617014
0.70001    0.70001         11.27   1.1155336
0.80001    0.80001          9.06   0.7905784
0.90001    0.90001         11.91   1.0422484
1.00001    1.00001         20.86   1.4099230    Discontinuity in ATFS
1.10001    1.10001         22.25   1.4797881
1.20001    1.20001         19.29   1.3295625
1.30001    1.30001         22.06   1.6701068
1.40001    1.40001         20.66   1.3877691
1.50001    1.50001         22.00   1.3193203
1.60001    1.60001         19.41   1.1540963
1.70001    1.70001         24.05   1.6870496
1.80001    1.80001         21.63   1.4109711
1.90001    1.90001         21.78   1.5225961
2.00001    2.00001         33.76   2.2924170    Discontinuity in ATFS
2.10001    2.10001         32.94   1.9074982
2.20001    2.20001         31.08   1.6842515
2.30001    2.30001         29.15   1.6531803
break to save space
6.30001    6.30001         71.27   3.2782064
6.40001    6.40001         70.86   2.5110270
6.50001    6.50001         73.37   2.4476191
6.60001    6.60001         69.88   2.5345005
6.70001    6.70001         70.29   2.8745591
6.80001    6.80001         68.14   2.5391488
6.90001    6.90001         79.08   2.6374031
7.00001    7.00001         82.50   2.8247285

[1] "step.h nr.loops"
[1] 1e-02 1e+03
        h.NormPois atfs.NormPois se.NormPois
6.89001    6.89001        78.842   0.9835133
6.90001    6.90001        78.587   0.9752885
6.91001    6.91001        78.702   0.9493621
6.92001    6.92001        79.278   0.9062284
6.93001    6.93001        80.901   0.9346886
```

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX D. GLOSSARY

These are medical terms from Dorland's Medical Dictionary

Cervical: Pertaining to the neck. Pertaining to the neck or cervix of any organ or structure.

Cervix: Latin word meaning neck. In anatomy, it is used for the neck and for any of a number of neck-like structures in the body.

Clinical: Pertaining to a clinic or to the bedside. Pertaining to or founded on actual observation and treatment of patients, as distinguished from theoretical or experimental.

Communicable Disease: A disease whose causative agents may pass or be carried from one person to another directly or indirectly. Modes of transmission include (1) direct contact with body excreta or discharges from an ulcer, open sore, or respiratory tract; (2) indirect contact with inanimate objects such as drinking glasses, toys, or bedclothing; and (3) vectors such as flies, mosquitoes, or other insects capable of spreading the disease. Called also Contagious Disease.

Conjunctiva: The delicate membrane lining the eyelids and covering the eyeball (ocular conjunctiva).

Dyspnea: Breathlessness or shortness of breath; labored or difficult breathing. It is a sign of a variety of disorders and is primarily an indication of inadequate ventilation or of insufficient amounts of oxygen in the circulating blood.

Edema: The accumulation of excess fluid in a body compartment; it may be in the cells (cellular edema), in the intercellular spaces within tissues (interstitial edema), or in potential spaces within the body. Edema may also be classified by location, such as pulmonary edema, cerebral edema. Edema can be caused by a variety of factors, such as an excess of hypotonic fluid (allowing movement of water into intracellular spaces), or decreased levels of plasma proteins (allowing passage of fluid out of the blood vessels into the tissue spaces). Other factors include poor lymphatic drainage; conditions that increase capillary pressure, such as excessive salt or fluid content of the blood, or heart failure; and conditions that increase capillary permeability, such as inflammation.

Emerging Infectious Disease: Is one that is endemic in a given population but that has begun increasing in frequency or developing resistance to drug therapy or other treatments. Infections that have newly appeared in a population or have existed but are rapidly increasing in incidence or geographic range

Endemic: Present or usually prevalent in a population or geographical area at all times, in contrast to Epidemic; said of a disease or infectious agent.

Epidemic: Occurring suddenly in numbers clearly in excess of normal expectancy, in contrast to Endemic or sporadic. The term is used especially of infectious disease but is also applied to any disease, injury, or other health related event occurring in such outbreaks.

Epiglottis: The lid-like cartilaginous structure overhanging the entrance to the larynx; the muscular action of swallowing closes the opening to the trachea by placing the larynx against the epiglottis, so that food and drink are prevented from entering the larynx and trachea and directed instead into the esophagus.

Etiology: Is the study or theory of the factors that cause disease and how disease is introduced to the host.

Hematogenous: Produced by or derived from the blood; disseminated through the bloodstream or by the circulation.

Hypotonic: Having abnormally reduced tonicity or tension. Having an osmotic pressure lower than that of the solution with which it is compared.

Infectious: Caused by or capable of being communicated by infection.

Infectious Disease: One due to organisms ranging in size from viruses to parasitic worms; it may be contagious in origin, result from nosocomial organisms, or be due to endogenous microflora from the nose and throat, skin, or bowel.

Larynx: The muscular and cartilaginous structure, lined with mucous membrane, situated at the top of the trachea and below the root of the tongue and the hyoid bone; it contains the vocal cords and is the source of the sound heard in speech. The larynx is part of the respiratory system; air passes through it traveling from the pharynx to the trachea on its way to the lungs and again returning to the exterior.

Lymph: Transparent, usually slightly yellow, often opalescent liquid found within the lymphatic vessels, collected from tissues in all parts of the body and returned to the blood via the lymphatic system. It is about 95 per cent water; the remainder consists of plasma proteins and other chemical substances contained in the blood plasma, but in slightly smaller percentage than in plasma. Its cellular component consists chiefly of lymphocytes.

Lymph node: Any of the accumulations of lymphoid tissue organized as definite lymphoid organs along the course of lymphatic vessels; they consist of an outer cortical and an inner medullary part. Lymph nodes are the main source of lymphocytes of the peripheral blood and, as part of the reticuloendothelial system, serve as a defense mechanism by removing noxious agents such as bacteria and toxins; they probably also play a role in antibody formation. Sometimes called, incorrectly, lymph gland. Called also lymph or lymphatic follicle and lymphatic nodule.

Morbid: Pertaining to, affected with, or inducing disease; diseased. unhealthy; unwholesome. Characterized by preoccupation with gloomy or unwholesome feelings or thoughts.

Morbidity: A diseased condition or state. The incidence or prevalence of a disease or of all diseases.

Nosocomial: Pertaining to or originating in a hospital.

Ocular: Pertaining to the eye. Called also ophthalmic and optic.

Oropharynx: The part of the pharynx between the soft palate and the upper edge of the epiglottis.

Osmotic pressure: The pressure required to stop osmosis through a semi permeable membrane between a solution and a pure solvent; it is proportional to the osmolality of the solution.

Pathology: Is the branch of medicine treating of the essential nature of disease, especially of the changes in body tissues and organs that cause or are caused by disease and the structural and functional manifestations of a disease.

Periorbita: The periosteum of the bones forming the orbit, or eye socket.

Periosteum: A specialized connective tissue covering all bones of the body, and possessing bone-forming potentialities. Periosteum also serves as a point of attachment for certain muscles.

Pharyngitis: Inflammation of the throat (pharynx); called also sore throat.

Pharynx: The musculomembranous cavity, about 12.5 cm (5 inches) long, behind the nasal cavities, mouth, and larynx, communicating with them and with the esophagus. The pharynx may be divided into three areas: the nasopharynx above; the oropharynx in the middle, behind the mouth; and the laryngopharynx below. The nasopharynx is connected with the nasal cavities and provides a passage for air during breathing. The oropharynx and laryngopharynx provide passageways for both air and food.

Preauricular: In front of the auricle of the ear.

Purulent: Pertaining to or consisting of pus, containing pus.

Reservoir: A storage place or cavity. An alternate or passive host or carrier that harbors pathogenic organisms or parasites without injury to itself and serves as a source from which other individuals can be infected.

Symptom: Any indication of disease perceived by the patient.

Syndrome: Is a combination of symptoms that either result from a single cause or occur together so commonly that they constitute a distinct clinical picture.

Syndromic: Occurring as a syndrome.

Vector: In epidemiology, is a carrier, especially an animal such as an arthropod that transfers an infective agent from one host to another.

Zoonosis: A disease of other animals that is transmissible to humans under natural conditions.

LIST OF REFERENCES

Arnon, S. S., Schechter, R., Inglesby, T.V., Henderson, D.A., Barlett, J.G., et al. (2001). Botulinum toxin as a biological weapon: Medical and public health management. *American Medical Association, 285*(8), 1059–1069, *Reprinted 285*(16), 2081.

Barns, S. M., Grow, C. C., Okinaka, R.T., Keim, P., and Kuske, C.R. (2005). Detection of diverse new *Francisella*-like bacteria in environmental samples. *American Society for Microbiology, 71*(9), 5494–5500.

BARDA (2010). BARDA funds drug development for biothreats, antibiotic resistance. *United States Department of Human & Health Services*. Retrieved November 15, 2010, from http://www.hhs.gov/news/press/2010pres/08/20100830a.html

Bhopal, R. S. (2002). *Concepts of epidemiology: An integrated introduction to the ideas, theory, principles and methods of epidemiology* [Adobe Digital Editions version]. New York, NY: Oxford University Press Inc. Retrieved from www.fk.uwks.ac.id/elib/Arsip/E.../Concepts%20of%20Epidemiology.pdf

Borio, L., Inglesby, T.V., Peters, C.J., Schmaljohn, A.L., Hughes, J.M., et al. (2002). Hemorrhagic fever viruses as biological weapons: Medical and public health management. *American Medical Association, 287*(18), 2391–2405.

CDC (2002). Tularemia – United States, 1990-2000. *Morbidity and Mortality Weekly Report, 51*(9), 181–184.

CDC (2010). Bioterrorism agents/diseases. Retrieved June 8, 2010, from http://emergency.cdc.gov/agent/agentlist-category.asp - a

Chu, M., Mead, P., and Sjöstedt, A. (2010). Tularemia. In D. L. Heymann (Ed.), *Control of Communicable Diseases Manual (CCDM18, CCDM19).* Retrieved December 15, 2010, from http://ccdm.unboundmedicine.com/ccdm/ub/view/Communicable-Diseases/460025/all/tularemia_

CIDRAP (2005). Tularemia agent found in DC air, but no cases seen. Retrieved November 5, 2010, from http://www.cidrap.umn.edu/cidrap/content/bt/tularemia/news/oct0305tularemia.html

CIDRAP (2010). Tularemia: Overview. *University of Minnesota*. Retrieved September 17, 2010, from http://www.cidrap.umn.edu/cidrap/content/bt/tularemia/biofacts/tularemiafactsheet.html [last update March 16, 2010]

Croddy, E. (2001). Tularemia, biological warfare, and the battle for Stalingrad (1942–1943). *Military Medicine, 166*(10), 837–838.

Cronquist, S. D. (2004). Tularemia: The disease and the weapon. *Dermatologic Clinics, 22*(3), 313–320.

Davis, C. J. (1999). Nuclear blindness: An overview of the biological weapons programs of the former Soviet Union and Iraq. *Emerging Infectious Diseases, Special Issue 5*(4), 509–512.

Dennis, D. T., Inglesby, T.V., Henderson, D.A., Bartlett, J.G., Ascher, M.S. et al. (2001). Tularemia as a biological weapon: Medical and public health management. *American Medical Association, 285*(21), 2763–2773.

Dorland's Medical Dictionary (2007). Elsevier. Retrieved from http://www.mercksource.com/pp/us/cns/cns_hl_dorlands_split.jsp?pg=/ppdocs/us/common/dorlands/dorland/misc/dmd-a-b-000.htm

Dudley, J. P. (2010). Tularemia: A case study in medical surveillance and bioterrorism preparedness. *Medical, Chemical, Biological, and Radiological Defense, 8*, 1–14. Retrieved from http://www.jmedcbr.org/issue_0801/Dudley/Dudley_09_10.html

Feldman, K. A., Enscore, R. E., Lathrop, S. L., Matyas, B. T., McGuill, M., et al. (2001). An outbreak of primary pneumonic tularemia on Martha's Vineyard. *New England Journal of Medicine, 345*(22), 1601–1606.

Foley, J. E., Nieto, N. C. (2010). Tularemia. *Veterinary Microbiology, 140*(3–4), 332–338.

Fricker, R. D., Jr. (2010a). *Introduction to Statistical Methods for Bioterrorism*. To be published by Cambridge University Press, in draft.

Fricker, R. D., Jr. (2010b). Some Methodological Issues in Biosurveillance (with discussion and rejoinder), *Statistics in Medicine*, to appear.

Fricker, R. D., Jr. (2010c). Biosurveillance: Detecting, Tracking, and Mitigating the Effects of Natural Disease and Bioterrorism, Encyclopedia of Operations Research and the Management Sciences, to appear.

Fricker, R. D., Jr. (2008). Syndromic Surveillance, in *Encyclopedia of Quantitative Risk Analysis and Assessment*, Melnick, E., and Everitt, B. (eds.), John Wiley & Sons Ltd., 1743–1752.

Fricker, R. D., Jr., & Rolka, H.R., (2006). Protecting against biological terrorism: Statistical issues in electronic biosurveillance. *Chance, 19*, 4–13.

Fricker, R. D., Jr., Hegler, B.L., and Dunfee, D.A., (2008). Comparing syndromic surveillance detection methods: EARS' versus a CUSUM-Based Methodology. *Statistics in Medicine, 27*, 3407–3429.

Fricker, R. D., Jr., Knitt, M.C., and Hu, C.X. (2008). Comparing directionally sensitive MCUSUM and MEWMA procedures with application to biosurveillance. *Quality Engineering, 20*, 478–494.

Grey, M. R., & Spaeth, K.R. (2006). *The bioterrorism sourcebook*. San Francisco, CA: McGraw-Hill.

Hagen, K. S. (2010). *Assessing the effectiveness of the early aberration reporting system (EARS) for early event detection of the H1N1 ("Swine Flu") virus*. Monterey, California: Naval Postgraduate School.

Han, S. W., Tsui, K-L., Ariyajunya, B., Kim, S.B. (2009). A comparison of CUSUM, EWMA, and temporal scan statistics for detection of increases in Poisson rates. *Quality and Reliability Engineering International, 26*(3), 279–289. doi: 10.1002/qre.1056

Hawkins, D. M. & Olwell, D. H. (1998). *Cumulative sum charts and charting for quality improvement*. New York, NY: Springer-Verlag, Inc.

Henderson, D. A., Inglesby, T. V., Bartlett, J. G., Ascher, M. S., Eitzen, E., et al. (1999). Smallpox as a biological weapon: Medical and public health management. *American Medical Association, 281*(22), 2127–2137.

Henning, K.J. (2003). What is syndromic surveillance? Syndromic surveillance: Reports from a national conference, 2003. *Morbidity and Mortality Weekly Report, 53* (Suppl), 7–11.

Hornick, R. (2001). Tularemia revisited. *New England Journal of Medicine, 345*(22), 1637–1639.

Inglesby, T. V., Dennis, D. T., Henderson, D. A., Berlett, J. G., Ascher, M. S., et al. (2000). Plague as a biological weapon: Medical and public health management. *American Medical Association, 283*(17), 2281–2290.

Inglesby, T. V., Henderson, D. A., Berlett, J. G., Ascher, M. S., Eitzen, E., et al. (1999). Anthrax as a biological weapon: Medical and public health management. *American Medical Association, 281*(18), 1735–1745, *283*(15), 1963.

Inglesby, T. V., O'Toole, T., Henderson, D. A., Berlett, J. G., Ascher, M. S., et al. (2002). Anthrax as a biological weapon, 2002: Update recommendations for management. *American Medical Association, 288*(17), 2236–2252, *288*(15), 1849.

Lorden, G. (1971). Procedures for reacting to a change in distribution. *Annal of Mathematical Statistics. 42*, 1897–1908.

MedicineNet (n.d.). Medical Terms/Lymph Node. Retrieve December 5, 2010, from http://www.medterms.com/script/main/art.asp?articlekey=4213

Morse, S. S. (1995). Factors in the emergence of infectious diseases. *Emerging Infectious Diseases, 1*(1), 7–15.

Page, E. S. (1954) Continuous inspections schemes. *Biometrika. 41*, 100–115.

Reintjes, R., Dedushaj, I., Gjini, A., Jorgensen, R. T., Cotter, B., et al. (2002). Tularemia outbreak investigation in Kosovo: Case control and environmental studies. *Emerging Infectious Diseases, 8*(1), 69–73.

Rossi, G., Lampugnani, L., & Marchi, M. (1999). An approximate CUSUM procedure for surveillance of health events. *Statistics in Medicine, 18*, 2111–2122.

Schoenbach, V. J. & Rosamond, W. D. (2000). *Understanding the fundamentals of epidemiology an evolving text* [Adobe Digital Editions version]. http://www.epidemiolog.net

Seiger, G. L. (1961). Ten uses of epidemiology. *Canadian Medical Association, 85*(18), 992–995.

Shmueli, G., & Burkom, H. (2010). Statistical challenges facing early outbreak detection in biosurveillance. *Technometrics, 52*(1), 39–51.

Sjöstedt, A. (2010). Tularemia. In D. L. Heymann (Ed.), *Control of Communicable Diseases Manual (CCDM18, CCDM19).* Retrieved December 15, 2010, from http://ccdm.unboundmedicine.com/ccdm/ub/view/Communicable-Diseases/460025/all/tularemia_

Trevisanato, S. I. (2007). The 'Hittite plague', an epidemic of tularemia and the first record of biological warfare. *Medical Hypotheses, 69*, 1371–1374.

Ugarte, M. D., Militino, A.F., & Arnholt, A.T. (2008). *Probability and statistics with R.* Boca Ratón, FL: CRC Press, Taylor & Francis Group, LLC.

Weinberg, A. N. & Branda, J. A. (2010). Case 31-2010: A 29-year-old woman with fever after a cat bite. *New England Journal of Medicine, 363*, 1560–1568.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Professor Ronald D. Fricker, Jr.
   Operations Research Department
   Naval Postgraduate School
   Monterey, California

4. Krista D. Hanni, PhD
   Monterey County Health Department
   Salinas, California